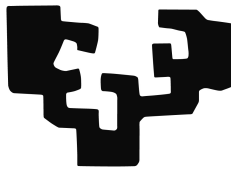


eman ta zabal zazu



EUSKAL HERRIKO UNIBERTSITATEA (UPV/EHU)
Hizkuntzaren eta Literaturaren Didaktika Saila

DOKTOREGO-TESIA

**Sentimenduen analisi automatikorantz:
oinarrizko baliabideen sorkuntza eta
hizkuntza maila ezberdinetako
balentzia-aldatzaileen identifikazioa**

Jon Alkorta Agirrezabala

Donostia, 2019

TESI ZUZENDARIAREN BAIMENA TESIA AURKEZTEKO	AUTORIZACIÓN DEL/LA DIRECTORA/A DE TESIS PARA SU PRESENTACIÓN
--	--

Zuzendariaren izen-abizenak /Nombre y apellidos del/la director/a:

Koldo Gojenola Gallettebeitia NA: 30571597C

Mikel Iruskiet Quintian NA: 30692223B

Tesiaren izenburua / Título de la tesis:

Sentimenduen analisi automatikorantz: oinarrizko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia-aldatzaileen identifikazioa

Doktorego programa / Programa de doctorado: **Hizkuntza Azterketa eta Prozesamendua**

Doktoregaiaren izen-abizenak / Nombre y apellidos del/la doctorando/a: **Jon Alkorta Agirrezabala**

Baimena ematen dut goian aipatzen den tesia aurkez dadin, horretarako baldintza guztiak betetzen baititu.

Autorizo la presentación de la tesis doctoral arriba indicada, dado que reúne las condiciones necesarias para su defensa.

Tokia eta data / Lugar y fecha:

Donostian, irailaren 25ean

Sin. / Fdo.: Tesiaren zuzendaria / El/La director/a de la tesis



Firmado por GOJENOLA
GALLETEBEITIA KOLDOBIKA -
30571597C el día 25/09/2019
con un certificado emitido por
AC FNMT Usuarios



Universidad Euskal Herriko
del País Vasco Unibertsitatea

TESI TUTOREAREN BAIMENA TESIA AURKEZTEKO

Nik, **Miren Igone Zabala Unzalu** doktoreak, tesi-tutorea naizen aldetik, baimena ematen dut tesi hau aurkez dadin, horretarako baldintza guztiak biltzen baititu:

Tesia: *Sentimenduen analisi automatikorantz: oinarrizko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia aldatzaileen identifikazioa*

Doktorego programa: Hizkuntza Azterketa eta Prozesamendua _____

Doktoregaia: Jon Alkorta Agirrezabala

Tessi zuzendaria: Koldo Gojenola Gallettebeitia eta Mikel Iruskieta Quintian

Lekua eta data: _Leioan, 2019ko irailaren 10ean

TESI TUTOREA

Iz.: Miren Igone Zabala Unzalu

DOKTOREGO PROGRAMAREN BATZORDE AKADEMIKOAREN BAIMENA

Bilbon Doktorego Programaren Batzorde Akademikoak 2019ko Irailaren 18an egindako bileran ondorengo izenburua duen doktorego tesia aurkeztearen alde azaltzea erabaki du: ***Sentimenduen analisi automatikorantz: oinarrizko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia – aldatzaileen identifikazioa*** izenburua duena.

Koldo Gojenola Galletebeitia eta **Mikel Iruskieta Quintian** jaun/andreek zuzendu eta **Jon Alkorta Agirrezabala** jaun/andreak aurkeztu du
Sail honen aurrean.

Bilbon, 2019ko Irailaren 18an

DOKTOREGO-PROGRAMAREN KOORDINATZAILEA

Sin.: Koldo Gojenola Galletebeitia

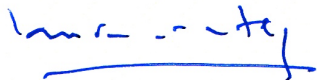
Firmado por GOJENOLA GALLETEBEITIA
KOLDOBIKA - 30571597C el día 18/09/2019
con un certificado emitido por AC FNMT
Usuarios

AUTORIZACIÓN DEL DEPARTAMENTO

El Consejo del Departamento de **Didáctica de la Lengua y la Literatura** en reunión celebrada el día **23 de septiembre de 2019** ha acordado dar la **conformidad** a la admisión a trámite de presentación de la Tesis Doctoral titulada: “Sentimenduen analisi automatikorantz: oinarrizko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia aldatzaileen identifikazioa” dirigida por los Dres. Kolde Gojenola Gallettebeitia y Mikel Iruskietia Quintian y presentada por D. **Jon Alkorta Agirrezabala** ante este Departamento.

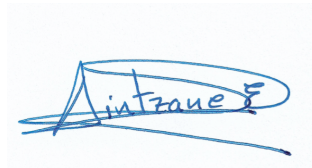
En Lejona a 23 de septiembre de 2019

VºBº DIRECTORA DEL DEPARTAMENTO



Fdo.: Laura Mintegi Lakarra

SECRETARIA DEL DEPARTAMENTO



Fdo.: Aintzane Etxebarria Lejarreta

DOKTORE GRADUAREN AKTA
DOKTOREGO TESIAREN DEFENTSAREN AKTA

Doktoregaiaren izen-deiturak: Jon Alkorta Agirrezabala
TESIAREN IZENBURUA: Sentimenduen analisi automatikorantz: oinarrizko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia-aldatzaileen identifikazioa

Goian aipatutako doktoregaiak tesia defendatu du UPV/EHUko Graduondoko Batzordeak tesia kalifikatzeko izendatutako epaimahaiaren aurrean, eta erantzun die hark aurkeztu dizkion eragozpen/galderei; defentsa eta erantzunok entzunda, epaimahaia bildu da datako egunean eta tesiari honako kalifikazio hau eman dio _____ (aho batez edo gehiengoaz):

BIKAIN / OSO ONGI / GAI / EZ GAI

Doktoregaiak hizkuntza hauek erabili ditu defentsan (bat baino gehiago erabili baditu, zehaztu hizkuntza bakoitzean defendatutako tesiaren ehunekoak):

Gaztelania: _____

Euskara: _____

Beste batzuk (zehaztu zeintzuk eta ehunekoak):

Lekua eta data: _____

EPAIMAHAIBURUA,

IDAZKARIA

Iz.: _____ dok.

Iz.: _____ dok.

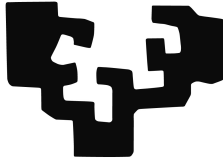
1. EPAIMAHAIKIDEA
Iz.: _____ dok

2. EPAIMAHAIKIDEA
Iz.: _____ dok

3. EPAIMAHAIKIDEA
Iz.: _____ dok

DOKTOREGAIA: Jon Alkorta

eman ta zabal zazu



EUSKAL HERRIKO UNIBERTSITATEA (UPV/EHU)

Hizkuntzaren eta Literaturaren Didaktika Hizkuntzaren eta Literaturaren
Didaktika Saila

**Sentimenduen analisi automatikorantz:
oinarrizko baliabideen sorkuntza eta
hizkuntza maila ezberdinetako
balentzia-aldatzaileen identifikazioa**

Jon Alkorta Agirrezabalak Koldo Gojenola
Galletebeitia eta Mikel Iruskietak Quintianen
zuzendaritzapean egindako tesiaren txostena,
Euskal Herriko Unibertsitatean Doktore titulu-
a eskuratzeko aurkeztua.

Aitari, Amari eta Josuri

Eskerrak

Lehenik eta behin, nire zuzendariak eskertu nahiko nituzke: Koldo eta Mikel, eman didaten laguntzagatik. Eskerrik asko beti nire atzean egoteagatik eta baita tesian zehar sortu diren zalantzetan laguntza eskaintzeagatik ere. Azkenik, eskerrik asko egin dizkidazuen iruzkinengatik (asko lagundu baitidate) eta ikerkuntzaren mundua nolakoa den erakusteagatik ere. Halaber, eskerrak Maxuxi eta Rodriri tesi-txostena hobetzen laguntzeagatik.

Eskerrik asko tesian lagundu didaten beste pertsoneri ere. Eskerrik asko Arantxari *Eustaggerrekin* izan ditudan une zailetan laguntzeagatik eta Intxari *includek* Latexen duen garrantzia erakusteagatik. Eskerrak Estherri Kanadatik Ixarako konexioa ahalbidetzeagatik, Kikeri emandako laguntzagatik eta Amaiari administrazioaren munduan laguntzeagatik. Eskerrik asko, halaber, Itziar Gonzalez-Diosi *Murritzapen Gramatikarekin* eta tesi-txostena vs. Latex borrokan laguntzeagatik. Eskerrik asko zerrenda honetan aipatu gabe gelditu diren baina lagundu didaten beste guztiei ere.

Eskertzekoa da tesia hasi eta amaitu arte inguruan mugitu denari ere: 318 bulegoari giro ona sortzeagatik, tupper txokoari askotariko gaiak (batzuetan, frikiak) aipatzeagatik, pintxo-poteetan eta egindako beste planetan atera zaretan guztioi eta, azkenik, estatistika klaseetan, Oierri izandako pazientziagatik eta *nukleo durori* estatistika ikasteko laguntza morala emateagatik.

También tengo que agradecer a Maite Taboada por su ayuda en la estancia de Vancouver y en los trabajos relacionados con el discurso y análisis de sentimientos.

Familia ere ezin da aipatu gabe utzi. Eskerrik asko aitari, amari eta Josuri

beti hor egoteagatik, eta laguntza behar izan dudanean laguntzeagatik.

Eskerrik asko kuadrillari ere, asteburuetan, musa eta *Comuniorekin* une bizi-ki ziragarriak emateagatik. Eskerrak baita Urkizu eta Gazteluko bazkariengatik eta Kanpezura eta Jakatik mendira egindako planengatik ere.

Bukatzeko eskerrik asko Amets eta Lierni pisukideei eta aipatzea ahaztu zaizkidan beste guztiei.

Mila-mila esker denoi!

Laburpena

Tesi-lan honetan, hizkuntzalaritza aplikatuaren ikuspegitik, euskarazko sentimenduen analisisian lehen urratsak egin dira. Bi helburu nagusi egon dira tesi-lanean. Alde batetik, sentimenduen analisisa egiteko oinarritzko baliabideak sortu ditugu euskararentzat. Zehatz esanda, Euskarazko Iritzi Corpusa, *Sentitegi* izeneko euskarazko sentimenduen lexikoa eta dokumentu mailako sentimenduen sailkatzailea garatu ditugu. Corpusak sei domeinutako 240 iritzi-testu biltzen ditu. Egitura Erretorikoaren Teoriaz (RST) baliatuta, corpuseko diskurtso-informazioa etiketatuta dago. Gainera, iritzi-testuen orientazio semantikoa ere etiketatuta dago. Sentimenduen lexikoari dagokionez, 1.237 hitzez osatuta dago eta bertako sarrerek -5 eta +5 arteko sentimendubalentzia dute. Sentimenduen lexikoa sortzeko itzulpen-metodologia zehatz bat jarraitu dugu. Azkenik, dokumentu mailako sentimenduen sailkatzailea ere garatu dugu. Tresnaren oinarrian aurretik aipatu dugun sentimenduen lexikoa dago.

Beste aldetik, sentimenduen analisiaren lanketa teoriko bat ere egin dugu. Sentimenduen sailkapena lexikoian oinarrituz egin nahi bada, hitzen sentimenduen balentzia jakitearekin ez da nahikoa, izan ere, testuetan badaude zenbait fenomeno hitz horien sentimenduen balentzia eragiten dutenak. Horiei testuinguruko balentzia-aldatzaileak deritze eta horiek euskaran nola agertzen diren landu dugu. Hizkuntza maila bakoitzeko balentzia-aldatzaile mota bat landu dugu: fonologian, bustidura adierazkorra; morfologian, hizkiak; sintaxian, ezeztapen-markak eta, azkenik, diskurtsoan, erlaziozko diskurtso-egiturak eta unitate zentrala. Emaitzek erakusten dutenez, balentzia-aldatzaileek hitzen edo sintagmen sentimenduen balentzia indartu

edo ahuldu egiten dute. Ahultze horren intentsitatearen arabera, sentimendu balentziaren zeinuan aldaketa gerta liteke, positiboa dena negatibo bilakatzuz edo alderantziz. Azkenik, kasu batzuetan, balentzia-aldatzaileak ez du eraginik sortzen.

Aurkibidea

Eskerrak	xv
Laburpena	xvii

SARRERA

1 Proiektuaren nondik norakoak	3
1.1 Motibazioa	3
1.2 Hipotesi orokorrak	6
1.3 Helburuak	8
1.4 Argitalpenak	9
1.5 Tesi-lanaren antolakuntza	12

AURREKARIAK

2 Sentimenduen analisiko baliabideak eta testuinguruko balentzia-aldatzaileak	17
2.1 Sentimenduen analisiaren atazak	18
2.1.1 Entitate eta aspektu mailako sentimenduen analisia . .	19
2.1.2 Esaldi mailako sentimenduen analisia	21
2.1.3 Dokumentu mailako sentimenduen sailkapena	24
2.1.4 Sentimenduen analisiko beste atazak	36

AURKIBIDEA

2.2	Sentimenduen analisirako baliabideen sorkuntza	43
2.2.1	Iritzi-testuen corpusak	43
2.2.2	Sentimenduen lexikoia	49
2.3	Balentzia-aldatazailerak	57
2.3.1	Testuingurua kontuan hartzen ez duten lanak	57
2.3.2	Berrikuntza: testuinguruko balentzia-aldatazailerak	60
2.3.3	Balentzia-aldatazaileretan oinarritzen diren lanak	64
2.4	Sentimenduen analisisa eta euskara	66
2.5	Laburpena	71

METODOLOGIA

3	Metodologiaren diseinua	75
3.1	Sentimenduen analisirako baliabide eta tresnen sorkuntza	78
3.1.1	Euskarazko Iritzi Corpora	78
3.1.2	<i>Sentitegi</i> izeneko euskarazko sentimenduen lexikoia- ren sorkuntza eta ebaluazioa	87
3.1.3	Lexikoian oinarritutako dokumentu mailako euskaraz- ko sentimenduen sailkatzailea	95
3.2	Balentzia-aldatazaileraren identifikazioa	100
3.2.1	Fonologia eta morfologia: bustidura adierazkorra eta hizkiak	100
3.2.2	Sintaxi maila: ezeztapen-markak	102
3.2.3	Diskurtsoa: erlazio diskurtso-egiturak, beren osagaiak eta unitate zentrala	109
3.3	Laburpena	117
4	Sentimenduen analisirako baliabideak	119
4.1	Euskarazko Iritzi Corpora	119
4.1.1	Euskarazko Iritzi Corpusaren ezaugarriak	119
4.1.2	Euskarazko Iritzi Corpusaren garapena	120
4.1.3	Euskarazko Iritzi Corpusaren baliagarritasunaren neur- keta	123

4.2	Euskarazko sentimenduen lexikoa	128
4.2.1	Euskarazko sentimenduen lexikoiaren ezaugarriak . . .	128
4.2.2	Lexikoiaren sorkuntza	130
4.2.3	Euskarazko sentimenduen lexikoiaren ebaluazioa	142
4.3	Euskarazko sentimenduen sailkatzailea	148
4.3.1	SO-CAL izeneko sentimenduen sailkatzailearen ezaugarriak	148
4.3.2	Sentimenduen sailkatzailearen arkitektura	151
4.3.3	Sentimenduen sailkatzailearen ebaluazioa	153
4.4	Laburpena	157

BALENTZIA ALDATZAILEAK HIZKUNTZA MAILA EZBERDINETAN

5	Balentzia-aldatzaileak	161
5.1	Maila fonologiko eta morfologikoa	161
5.1.1	Balentzia-aldatzaileen sailkapena	161
5.1.2	Hainbat balentzia-aldatzaile morfologiko dituzten hitzak	165
5.1.3	Balentzia-aldatzaileak orientazio semantiko ezberdinetako hitzetan	167
5.1.4	Maila fonologikoa eta morfologikoaren garrantzia euskaran	169
5.2	Sintaxi maila	172
5.2.1	Ezeztapenezko balentzia-aldatzaileak	172
5.2.2	Trukaketa- eta desplazamendu-ezeztapena	175
5.2.3	Ezeztapen-markak eta beren irismena identifikatzeko erregelak	176
5.3	Diskurtso maila	184
5.3.1	Balentzia-aldatzaileak: nukleartasuna eta unitate zentrala	184
5.3.2	Egitura Erretorikoaren Teoria (RST)	194
5.4	Laburpena	199

AURKIBIDEA

ONDORIOAK

6	Ekarpenak, mugak eta etorkizuneko lanak	203
6.1	Ekarpenak	203
6.1.1	Sentimenduen analisirako baliabideak eta tresnak	204
6.1.2	Euskarazko balentzia-aldatzaileen azterketa	204
6.2	Lanaren mugak	206
6.3	Etorkizuneko lanak	207

	Bibliografia	208
--	---------------------	------------

	Terminologia eta laburdurak	229
--	------------------------------------	------------

ERANSKINA

A	Murriztapen Gramatikako erregelak	233
----------	--	------------

Irudien zerrenda

2.1	<i>Pointwise mutual information</i> , PMI.	29
2.2	Aspektuan oinarritutako iritzi-testuen laburpenaren emaitza egituratua (Cambria <i>et al.</i> , 2017).	37
2.3	Iritziaren kalitatearen neurketa Amazon webgunean.	41
2.4	ML-SentiCon lexikoiaren ingelesezko bertsioaren zati bat (Cruz <i>et al.</i> , 2014).	67
3.1	Tesi-lanaren metodologia.	76
3.2	<i>Kritiken Hemerotekaren</i> webgunea.	79
3.3	Euskarazko aditzetan dagoen lehen pertsonaren erabileraren neurketa.	81
3.4	EGU04 iritzi-testuaren diskurtso-egituraren anotazioa.	83
3.5	<i>Ondorioz</i> hitza eta bere testuinguruak.	91
3.6	SO-CAL tresnaren ingelesezko eta euskarazko bertsioen egituraren alderaketa.	97
3.7	Ezeztapen-markak eta beren irismen-eremua identifikatzeko erregelen adibide bat <i>Murritzapen Gramatika</i> (Karlsson <i>et al.</i> , 1995) ingurunean.	107
3.8	<i>Murritzapen Gramatika</i> n (Karlsson <i>et al.</i> , 1995) sortutako erregelek esleitutako ! etiketa.	108
3.9	Erregelak ebaluatzeko etiketatzaileak hitzei esleitutako etiketak.	109
3.10	SENTFAR-01 testuaren RST-zuhaitza.	111
3.11	Bi pertsonak egin beharreko unitate zentralaren aukeraketa EGU01 iritzi-testuan.	113
4.1	Iritzi-testu hirueledun baten adibidea.	121
4.2	Euskarazko iritzi-testu labur baten adibidea.	122

IRUDIEN ZERRENDA

4.3	Gaztelaniazko lexikoiko hitzei euskarazko ordaina emateko gauzatutako urratsak.	137
4.4	SO-CAL sentimenduen sailkatzailearen euskarazko bertsioa.	151
4.5	EGU40 iritzi-testua.	152
4.6	EGU40 iritzi-testua lematizatuta eta hitzei sentimendu-balentziak esleituta.	153
4.7	EGU40 iritzi-testuko hitzei sentimenduen sailkatzailearen erregelak aplikatuta.	154
4.8	EGU40 iritzi-testuaren sentimendu-balentzia.	154
5.1	Ezeztapen-marken eragin ezberdinak sentimendu-balentzian.	172
5.2	Ixa taldearen analisi-katea (Oronoz, 2009).	180
5.3	Ixa taldeko analisi-katearen emaitza.	180
5.4	Testuek eta erlaziozko diskurtso-egiturek beren artean dituzten orientazio semantikoaren adostasun eta desadostasunak.	187
5.5	Ebaluazioa-Interpretazioa eta Ebidentzia-Justifikazioa erlaziozko diskurtso-egituren instantziak.	189
5.6	Kausa-Ondorioa-Helburua eta Antitesia-Kontzetsioa erlaziozko diskurtso-egitura taldeen instantziak.	190
5.7	EGU06 iritzi-testua RSTz etiketatuta.	195
5.8	Euskarazko balentzia-aldatzaileak hizkuntza maila ezberdinetan.	199

Taulen zerrenda

2.1	Esaldien subjektibitatearen sailkapena egiteko teknika ezberdinak.	22
2.2	Esaldi mailako sentimenduen sailkapena egiteko zenbait teknika.	23
2.3	Dokumentu mailako sentimenduen sailkapena egiteko erabilitako zenbait teknika.	27
2.4	Dokumentu mailako sentimenduen sailkapen gainbegiratu gaberako erregela sintaktikoak (Turney, 2002).	28
2.5	Sentimenduen estimazio-iragarpena egiteko teknika ezberdinak.	32
2.6	Esaldi mailako hizkuntza arteko sentimenduen analisirako zenbait teknika.	34
2.7	Dokumentu mailan, hizkuntza arteko sentimenduen analisisa egiteko zenbait teknika.	35
2.8	Konparaziozko iritziak lantzeko zenbait teknika.	38
2.9	Iritzi-bilaketa egiteko aipatutako teknikak.	39
2.10	Iritzien kalitatea neurtzeko zenbait teknika.	41
2.11	Sentimenduen analisirako sortu diren zenbait corpus.	45
2.12	Hiztegian oinarrituz sentimendu lexikoa sortzeko zenbait teknika.	51
2.13	Corpusean oinarrituta sortutako zenbait sentimenduen lexikoi.	53
2.14	Hatzivassiloglou eta McKeownen (1997) laneko hazi-hitzak. . .	53
2.15	Hatzivassiloglou eta McKeownen (1997) sentimenduen lexikoa zati bat.	54
2.16	Hitzen orientazio semantikoa esleitzeko testuingurua kontuan hartzen eta hartzen ez duten lanak.	58
2.17	Sinonimo hurkoen sailkapena beren ezberdintasunetan oinarrituta (Edmonds eta Hirst, 2002, 5. orr.).	58

2.18	Euskara eta sentimendua uztartzen dituzten zenbait lan. . . .	66
3.1	Datu-basearen antolaketaren bi adibide.	80
3.2	Bi anotatzailek (A1 eta A2) anotatutako iritzi-testuen kopurua.	82
3.3	Bi etiketatzaileen arteko adostasun-maila iritzi-testuak RST hurbilpenaz etiketatzean (eskuzko ebaluazioa).	83
3.4	Etiketatzailen arteko adostasunari buruz tresna automati- koak egindako ebaluazio kualitatiboa.	84
3.5	Erlaziozko diskurtso-egitura baten orientazio semantikoaren esleiketa.	85
3.6	Etiketatutako erlaziozko diskurtso-egiturak.	86
3.7	Erlaziozko diskurtso-egituren sentimendu balentzia anotazioa- ri dagokion bi anotatzaileen arteko kontingentzia-taula. . . .	86
3.8	SO-CAL tresnaren (Taboada <i>et al.</i> , 2011) gaztelaniazko ber- tsioaren lexikoia zati bat.	89
3.9	SO-CAL tresnaren ingelesezko bertsioaren lexikoia zati bat.	90
3.10	Itzulpen-prozesuaren azalpeneko adibideak.	91
3.11	400 hitzeko zerrendaren zati bat etiketatzaile batek bertako hitzei sentimendu-balentzia esleituta.	95
3.12	Ingelesezko eta euskarazko sentimenduen lexikoien zati bat. . .	98
3.13	Euskarazko sentimenduen sailkatzailearen ebaluazioa.	99
3.14	Hizkiei buruz bildutako informazioa.	100
3.15	Corpuseko hitzen zerrendaren zati bat maiztasunean oinarrituta.	101
3.16	Corpusean agertutako hizkien eta bustidura adierazkoren az- terketaren lagin bat	101
3.17	Ezeztapen-markak aztertze sortutako azpicorpusaren zati bat.	103
3.18	Sentimendu-balentzian eragin ezberdinak dituzten ezeztape- nak identifikatzeko proposatu diren erregeletako batzuk. . . .	106
3.19	Bi pertsonen arteko adostasun-maila 78 iritzi-testuetan unita- te zentrala aukeratzerakoan.	114
3.20	<i>Analhitza</i> tresnak (Otegi <i>et al.</i> , 2017) eguraldiaren domeinuko hitzei egindako gramatika-kategoriaren sailkapena.	114
3.21	Eguraldiaren domeinuko unitate zentraleko hitzen zerrenda. .	115
3.22	Eguraldiaren domeinuan, gramatika-kategoria bakoitzeko sentimendu- balentziadun hitzen kopurua.	116

4.1	Sentimenduen analisirako eskura dauden iritzi-testuen corpusak.	120
4.2	Lehen pertsonaren eta adjektiboaren agerpena neurtzeko erabilitako corpusak.	124
4.3	Lehen pertsonaren agerpena lau corpus ezberdinetan.	126
4.4	Adjektiboen agerpena corpus objekibo eta subjektiboetan. . .	127
4.5	Euskarazko sentimenduen lexikoiaren bi bertsioen ezaugarriak.	128
4.6	Lexikoi paraleloaren adibide batzuk.	129
4.7	Euskarazko ordain posible bat baino gehiago dituzten adibideak.	133
4.8	Hitz polisemikoen tratamendua.	134
4.9	Itzulpen-prozesuaren koherentzia erakusten duten adibideak. .	134
4.10	Bi etiketatzaileraren arteko korrelazioaren kalkulua, etiketatu gabeko hitzak aintzat hartu gabe (Pearson 1).	143
4.11	Bi etiketatzaileraren arteko korrelazioaren kalkulua, etiketatu gabeko hitzak ere aintzat hartuz (Pearson 2).	144
4.12	Bi anotatzaileen arteko Pearson korrelazioaren neurketa. . . .	144
4.13	Bi anotatzaileen arteko kontigentzia-taula.	145
4.14	Euskarazko sentimenduen lexikoa (V2.0) eta urre-patroiaren arteko Pearson korrelazioaren neurketa.	146
4.15	Euskarazko sentimenduen lexikoa (V2.0) eta urre-patroiaren arteko kontigentzia-taula.	147
4.16	SO-CAL tresnako lexikoiaren zenbait sarrera (Taboada <i>et al.</i> , 2011).	149
4.17	SO-CAL tresnako zenbait intentsifikatzaile (Taboada <i>et al.</i> , 2011).	149
4.18	Sentimenduen sailkatzailearen ebaluazioaren emaitzak.	155
4.19	Sentimenduen sailkatzailearen ebaluazioaren emaitzak iritzi-testuen orientazio semantikoa aintzat hartuta.	155
4.20	Sentimenduen sailkatzailearen ebaluazioaren emaitzak domeinuen arabera.	156
5.1	Euskarazko Iritzi Corpusetik (Alkorta <i>et al.</i> , 2016) lortutako balentzia-aldatzaile morfologiko eta fonologikoak.	162
5.2	Balentzia-aldatzaile morfologiko eta fonologikoak sentimendubalentzian duten eraginaren arabera sailkatuta.	163

5.3	Euskarazko Iritzi Corpusetik lortutako hizkien sailkapen semantikoa.	164
5.4	Hitz batean hainbat hizki agertzeko moduak eta horien eragina hitzaren balentzian.	166
5.5	Balentzia aldatzen duten hizkiek sentimendu-balentzia ezberdineko hitzetan duten eragina aztertzeke taula.	167
5.6	Sentimendu-balentzian eragin ezberdinak dituzten ezeztape-nak identifikatzeko proposatu diren erregeletako batzuk. . . .	177
5.7	Emaitzak ezeztapen-markek sentimendu-balentzian eragiten dutenaren ikuspegitik.	181
5.8	Emaitzak ezeztapenari loturiko elementu-motaren ikuspegitik.	182
5.9	Orientazio semantikoaren adostasuna erlaziozko diskurtso-egituren eta nukleoaren/satelitearen artean.	185
5.10	Orientazio semantikoaren adostasuna erlaziozko diskurtso-egituren eta lehen osagaiaren/azken osagaiare artean.	186
5.11	Unitate zentraletan bereizgarri diren ezaugarriak domeinuen eta gramatika-kategorien arabera.	192
5.12	Euskarazko sentimenduen sailkatzaileak sentimendu-balentzia esleitutako unitate zentraleko hitzak gramatika-kategoriaren arabera sailkatuta.	193
5.13	Sentimendu-balentziadun hitzen agerpena.	193

SARRERA

Proiektuaren nondik norakoak

1.1 Motibazioa

Iritziak giza jardueraren oinarrian daude eta gure portaeran eragin handia dute (Liu, 2012). Errealitatearen gure uste eta pertzepzioak nahiz egiten ditugun aukeraketak hein handi batean munduak ikusi eta ebaluatzen duenaz baldintzatuta daude. Egoera horren ondorioz, erabaki bat hartu behar dugunean, besteen iritziak bilatu ohi ditugu. Norbanakoetan eta erakundeetan gertatzen den egoera da.

Iritziaren inguruan ere badira zenbait kontzeptu subjektibitatearekin lotura dutenak. Besteen artean, sentimendua, ebaluazioa, jarrerak eta emozioak aipa daitezke eta horiek guztiak, iritziaz gain, sentimenduen analisia edo iritzi-meatzaritzaren deritzon arloaren ikergaiak dira.

Liuk (2012) gogorarazten duenez, arlo horren hasiera eta hazkundera erabat loturik daude webetako sare sozialen jaiotzarekin. Webetako sare sozialak era askotakoak dira: iruzkinak, foroetako eztabaidak, blogak, mikroblogak eta Twitter. Aplikazio horiek guztiek gizakiaren historian lehen aldiz digitalki jasotako iritzi bolumen handi bat edukitzea ahalbidetu dute. Hori horrela izanik, 2000tik aurrera hizkuntza naturalaren prozesamenduan hazkunderik handiena eta aktiboena bilakatu den arloa da. Gainera, datu-meatzaritzan, web-meatzaritzan eta testu-meatzaritzan ere erabiltzen da.

Oro har, enpresan eta gizartean garrantzizkoa bilakatu da sentimendu-analisia, eta horrek arloa konputazio-zientzietatik kudeaketa-zientzietara eta gizarte-

zientzietara zabaldu du. Bere inguruan, industria garatu da eta enpresek zerbitzuak ere sortu dituzte arloa lantzeko. Beraz, sentimendu-analisiak enpresa edo gizarte eremu askotan aplikazioa aurkitu duela esan daiteke.

Ikusi dugun moduan, nazioartean sentimenduen analisiak garapen handia izan du. Egun, munduko *lingua franca* ingelesa da eta sentimendu-analisiko lan gehienak hizkuntza horretan egin dira. Gurera etorrira, ordea, gutxi dira euskara hartuta egin diren lanak.

Sentimendu-analisia oso arlo zabala da. Ataza batzuek lotura handiagoa dute informatikarekin eta beste batzuek, berriz, hizkuntzalaritzarekin. Gainera, ataza batzuetan bi hurbilpen egoten dira: hizkuntzaren ezagutzan oinarritutakoa eta metodo estatistikoetan oinarritutakoa.

Hizkuntzalaritzaren ikuspegitik, lehen urratsa da zein hitzek duen iritzia edo, modu zabalago batean, informazio subjektiboa den jakitea. Hurrengo urrartsean, ordea, hitz horien informazio subjektiboan eragiten duten hizkuntzari lotutako fenomenoak bilatu nahi dira. Azter ditzagun hurrengo adibideak.

- (1) Pelikula hori [gustatu₊]₊ zait.
- (2) Pelikula hori [asko gustatu₊]₊₊ zait.
- (3) Pelikula hori ez zait [gustatu₊]₋.
- (4) Pelikula hori [gustatuko₊ litzaidake]₀ kontakizuna biziagoa izango babiliz.

Goian, pelikula bati buruzko zenbait iritzi agertzen dira. Denak antzekoak dira, baina ez berdinak. (1) adibidean, esaterako, pelikula gustukoa izan duela esaten da; beraz, esaldiaren balorazioa positiboa da. (2) adibidean ere esaldiaren balorazioa positiboa da, baina *asko* intentsifikatzailea¹ agertzen da eta, horregatik, esaldi honek aurreko esaldiak baino iritzi positiboagoa du. (3) adibidean, ordea, esaldian iritzia adierazten duen hitz bat, positiboa gainera, egon arren, esaldiaren balorazioa negatiboa da. Azkenik, (4) adibidean ere, iritzia duen hitz bakarrak balorazio positiboa egiten du, baina esaldiak

¹Intentsifikatzaileak adberbioak edo adjektiboak izan ohi dira. Berezko eduki semantiko gutxi dute, baina aldatzen duen hitzaren edo esaldiaren esanahia intentsifikatzen du.

ez du iritzirik adierazten, erreala ez den egoera batean dagoelako iritzia².

Adibide horietan ikusten denez, esaldi guztietan iritzia adierazten duen hitz bat (*gustatu*) egon arren, esaldi osoek adierazten duten iritzia ezberdina da. Tartean, hitzen arteko interakzioak egon dira eta fenomeno linguistiko batzuek iritzia adierazten duen hitzaren balorazioa aldatu egin dute.

Hain zuzen ere, aipatu ditugun bi alderdi horiek dira gure lanaren motibazioak. Alde batetik, euskararentzat sentimendu-analisiaren tresnak eta baliabideak sortu nahi ditugu, gaur egun, euskarazko testuen eduki subjektiboa erauzi ahal izateko. Beste aldetik, berriz, iritzia edo informazio subjektiboa duten hitzetan eragiten duten fenomeno linguistikoak bilatu nahi ditugu, fonologian, sintaxian eta diskurtsoan.

Gure lehen helburua sentimendu-analisiaren arloan ikerketa egiteko oinarriko tresna eta baliabideak sortzea da. Zehazki, euskarazko iritzi-testuen corpusa eta euskarazko sentimendu lexikoa sortu nahi ditugu.

Izan ere, gutxi dira euskaraz dauden iritzi-testuen corpusak eta sentimendu lexikoiak, eta daudenak, berriz, ez dira baliagarriak gure helburuak betetzeko; izan ere, ez dituzte hizkuntzaren ezaugarriak aintzat hartzen eta hori ez dator bat gure ikusmoldearekin. Iritzi-testuen kasuan, txioak eta zerbitzuak (produktuen salerosketa webguneek eta hotelen web konparatzailea, adibidez) eskaintzen dituzten webguneetako iritziak biltzen dituzten corpusak edota datu-baseak badaude baina guri ez zaizkigu baliagarri testuak moztzak direlako eta ondorioz, diskurtsorako zenbait elementu aztertzeke aukera zailduko ligukeelako. Lexikoiei dagokienez, berriz, hitzen orientazio semantiko (positiboa edo negatiboa) adierazi arren, ez dute intentsitatea adierazten (hau da, *ondo* eta *bikain* positiboak dira, baina intentsitatean duten ezberdintasuna ez da adierazten) eta horregatik, ez zaizkigu baliagarriak. (Stone eta Hunt, 1963) sentimendu lexikoa, esaterako, tankera horretakoa da.

Gure bigarren helburua aurreko erronkan sortutako sentimendu lexikoiko hitzen balentzian eragiten duten hizkuntza-fenomenoak bilatzea eta beren eragina neurtzea da.

²Adibidean ematen den iritzia egoera errealean ez dagoela diogu, iritzia emateko erabiltzen den aditza ez dagoelako ez iraganean, ezta orainaldian ere.

(1), (2), (3) eta (4) adibideetan ikusi dugun moduan, lexikoian oinarritzen den sentimenduen sailkatzaile batean, lexikoia bere baitan bakarrik ez da baliagarria. Informazio linguistiko gehiago behar da sentimenduen sailkapen on bat egiteko. Horregatik, Polanyi eta Zaenenen (2006) lanean oinarrituz, euskaran dauden hizkuntza maila ezberdinetako testuinguruko balentzia-aldatzaileak³ identifikatu nahi ditugu.

Gure hirugarren eta azken helburua dokumentu mailako euskarazko sentimenduen sailkatzaile bat sortzea da. Euskaran lan gutxi daude sentimenduen sailkapena egiten dutenak. Horietako bat EliXa (San Vicente *et al.*, 2015) da eta aspektu mailako sentimenduen sailkapena egiten du⁴. Gure erronka euskarari hizkuntza-prozesamendurako beste baliabide bat ematea da eta, horretarako, lexikoian oinarritutako dokumentu mailako sentimenduen sailkatzailea sortu nahi dugu. Esan behar da erronka hau betetze-ko beharrezkoa dela aurretik aipatutako erronkak betetzea: sentimenduen lexikoi bat sortu behar da sentimenduen sailkatzailea inplementatzeko eta balentzia-aldatzaileak landu behar dira lexikoiko hitzetan eragiten duten al-daketa neurtzeko.

Laburbilduz, tesi honek euskara eta sentimenduen analisia uztartzen ditu. Euskara hizkuntzat hartzen duten lan gutxi egin dira sentimenduen analisian eta gehienak sentimendu lexikoia sortzera mugatzen dira. Guk euskarari baliabide eta tresnak eman nahi dizkiogu, baita euskarak arlo honetan dituen berezitasunak ikertu ere.

1.2 Hipotesi orokorrak

Aurreko atalean, sentimenduen analisia arloaren motibazioa aipatu dugu, baita berak gizartean duen erabilgarritasuna ere. Atal honetan, tesiaren irismena zehaztuko dugu. Lau hipotesi orokor ditu lan honek.

³Testuinguruko balentzia-aldatzaileak (*contextual valence shifters*, ingelesez) hitzen edota esaldien sentimendu balentzian aldatetak eragiten dituzten fenomenoak dira. Aldaketa hori balentziaren indartze bat edo ahultze bat izan daiteke.

⁴Aspektu mailako sentimenduen analisiak aztergai den entitateari loturiko aspektuak identifikatu (entitatea Londres bada, aspektuak ekonomia, turismoa etab. dira) eta horiei buruzko iritzia positiboa edo negatiboa den zehaztu nahi du.

1. ikerketa-hipotesia: *itzulpen bidez lortutako euskarazko sentimenduen lexikoiaaren kalitatea corpusetik edo datu-base lexikaletatik sor daitezkeenak bezain baliagarria da.*

Sentimenduen lexikoiek testuetako hitzei beren orientazio semantikoa⁵ eta sentimenduen balentzia esleitzen diete eta hori testu baten orientazio semantikoa kalkulatzeko lehen urratsa da. Lexikoak sortzeko, ohiko bi hurbilpen daude: i) corpus bidez eta ii) datu-base lexikalen bidez.

Guk hautatutako bidea aipaturiko hurbilpenen nahasketa da. Oinarri bezala bi datu-base lexikal, hots, gaztelaniazko eta ingelesezko bi sentimenduen lexikoi erabiliko ditugu eta horiek itzuli egingo ditugu Euskarazko Iritzi Corpusak ematen duen informazioa aberastuz.

Gure ustez, jada sorturik dagoen sentimenduen lexikoa euskaratzea aurreko bi hurbilpenak (corpus bidezkoa eta datu-base lexikal bidezkoa) bezain baliagarria da eta kalitatean ez legoke ezberdintasun handirik. Izan ere, itzulpenaren metodologia aproposa bada, bidean ez da informaziorik galduko edo eraldatuko eta hasierako kalitatea mantenduko luke. Gainera, iritzi-testuen corpora lagun izanda, sentimendua duten hitzen orientazio semantikoan nahiz sentimenduen balentzian okerreko esleipenak egitea eragotziko litzateke.

2. ikerketa-hipotesia: *hizkuntza maila guztietan (fonologia, sintaxia eta diskurtsoa) zenbait fenomeno linguistiko daude orientazio semantikoa duten hitzetan (eta esaldietan) eragiten dutenak.*

Eragin hori beren sentimenduen balentzia indartzea edo ahultzea izan daiteke. Batzuetan fenomeno linguistikoak ez du balentzian aldaketarik eragiten.

Gure ustez, fenomeno linguistiko horiek gramatika maila ezberdinetan ager daitezke: esateko moduan (hau da, fonologian), hizkien bidez, sintaxiko fenomeno ezberdinen bidez edota baita diskurtsoaren bidez ere. Lexikoian oinarritzen den dokumentu mailako sentimenduen sailkatzaile bati mota honetako informazioa gehitzea beharrezkoa dela uste dugu, bestela sailkatzailearen emaitzak ez baitira nahi bezain zehatzak izango.

⁵Orientazio semantikoa hitz batek duen informazio subjektiboa da eta positiboa edo negatiboa izan daiteke.

3. ikerketa-hipotesia: *diskurtso-egituran badaude osagaiak EDUen*⁶

Gure ustez, EDU edo erlaziozko diskurtso-egiturak orientazio semantiko positiboa edo negatiboa edukitzea ez da ausazko gertaera bat. Hau da, elementu horiek RST-zuhaitzean⁷ dituzten guneak horretan eragina dutela uste dugu. Zehazki esanda, diskurtso-egiturako zenbait faktorek edo osagaik EDU eta erlaziozko diskurtso-egiturako orientazio semantikoa baldintzatzen dutela da gure hipotesia.

4. ikerketa-hipotesia: *testu baten domeinuak unitate zentralak ezaugarri jakin batzuk izatea eragiten du, bai bertan agertzen diren hitzen gramatika-kategoriari dagokionez, bai hitz horiek orientazio semantikoa edukitzea edo ez edukitzearekin ere.*

Nahiz eta unitate zentrala testu guztietan dagoen, bere ezaugarriak domeinuaren arabera ezberdinak direla uste dugu. Eta horrek unitate zentraletan orientazio semantikodun hitzak agertzeko moduan eragiten duela ere pentsatzen dugu. Testuen sentimenduen analisia egiterakoan, beharrezkoa domeinua zein den jakitea, horrek arrasto batzuk ematen baititu sentimendu balentziadun hitzak non agertzen diren jakiteko.

1.3 Helburuak

Aurreko ataletan ikerketa-lan honen testuingurua eta gure ikerketa gidatzeko hipotesi orokorrak eztabaidatu ditugu. Atal honetan, berriz, gure helburu zehatzak adieraziko ditugu gure hipotesiak sostengatzen dituzten egiaztatze-ko.

- **1. helburua:** Sentimenduen analisia lantzeko oinarrizko tresnak eta baliabideak sortzea.

⁶Oinarrizko Diskurtso Unitatea (*Elementary Discourse Unit*, EDU) *diskurtso-egiturako unitaterik txikiena da.* eta erlaziozko diskurtso-egituren orientazio semantikoa duten bezalako izatea eragiten dutenak.

⁷Egitura Erretorikoaren Teoria (*Rhetorical Structure Theory*, RST) testu-sorkuntza automatikorako sortutako teoria da. RSTn testuaren antolakuntza testu-zatien artean daude erlaziozko diskurtso-egitura ezberdinen bidez garatzen da eta koherentzia testuaren egiturak duen hierarkiaren bidez gauzatzen da.

- **1.1 helburua:** euskarazko iritzi-testuak biltzen dituen corpusa sortzea, testuen orientazio semantikoari dagokionez orekatua eta sintaxiari dagokionez aberatsa.
- **1.2 helburua:** euskarazko sentimenduen lexikoa sortzea, hitzen orientazio semantikoa adierazten duena eta horretarako hitzei zenbakizko balio bat esleitzen diena.
- **1.3 helburua:** dokumentu mailako euskarazko sentimenduen sailkatzaile bat sortzea, oinarritzat sentimenduen lexikoa izango duena.
- **2. helburua:** hitzen sentimenduen balentzian eragiten duten euskarazko testuinguruko balentzia-aldatzaileak identifikatzea eta beren eragina neurtzea.
 - **2.1 helburua:** fonologiako balentzia-aldatzaileak identifikatzea eta beren eragina neurtzea.
 - **2.2 helburua:** morfologiako balentzia-aldatzaileak identifikatzea eta beren eragina neurtzea.
 - **2.3 helburua:** sintaxian, ezeztapen-markek hitzaren edo esaldia-
ren sentimendu balentzian nola eragiten duten neurtzea.
 - **2.4 helburua:** diskurtsoan, erlaziozko diskurtso-egituretan nukleartasunaren eta iritzi-testuetan unitate zentralaren eragina neurtzea orientazio semantikoari dagokionez.
 - **2.5 helburua:** iritzi-testuetan, erlaziozko diskurtso-egiturek testuaren posizio jakin batean agertzeko joera duten aztertzea.
 - **2.6 helburua:** iritzi-testuetako unitate zentralen azterketa: hitzen gramatika-kategoriaren azterketa eta orientazio semantikoa duten hitzen banaketaren lanketa.

1.4 Argitalpenak

Tesi-lan honetan zehar garatutako zenbait lan aldizkari eta kongresu ezberdinetan aurkeztu dira. Atal honetan, argitalpen horiek zerrendatu ditugu tesi-laneko gaiaren arabera sailkatuta.

- **Corpusa**

- Alkorta J., Gojenola K. eta Iruskieta M. Creating and evaluating a polarity-balanced corpus for basque sentiment analysis. In *Fourth International Workshop on Discourse Analysis*, 54-58, Santiago de Compostela, Spain, 2016a. ISBN 978-84-608-9305-9.

- **Sentimendu-hitzak eta sentimendu-lexikoia**

- Alkorta J., Gojenola J. eta Iruskieta M. SentiTegi: Semi-manually Created Semantic Oriented Basque Lexicon for Sentiment Analysis. *Computación y Sistemas* 22(4): 1295-1306, 2018b, ISSN 2007-9737, DOI 10.13053/CyS-22-4-3075
- Alkorta J., Gojenola J. eta Iruskieta M. Sentimenduak deskribatzeko hurbilpen teoriko konputazionala euskaraz. *UPV/EHUko I. Doktorego Jardunaldiak. Elkarrekin ikertuz: 245-246*, 2017a, Euskal Herriko Unibertsitateko Argitalpen Zerbitzua, ISBN 978-84-9082-619-5.
- Alkorta J., Gojenola J., Iruskieta M. eta Taboada M. Using lexical level information in discourse structures for Basque sentiment analysis. In *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, 39-47, Santiago de Compostela, Spain, 2017b. ISBN 978-1-945626-78-4.

- **Ezeztapena**

- Alkorta J., Gojenola J. eta Iruskieta M. Saying no but meaning yes: negation and sentiment analysis in Basque. In Balahur A., Mohammad S. M., Hoste V., eta Klinger, R., editoreak, *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 85-90, Brussels, Belgium, 2018a. The Association for Computational Linguistic. ISBN 978-1-948087-80-3.

- **Diskurtsoa**

-
- Alkorta J., Gojenola J., Iruskieta M. eta Perez A. El uso de la información de la estructura retórica en el análisis de sentimiento. In Martínez Barco P., Navarro Colorado B., Vázquez Pérez S., eta Romá Ferri M.T., editoreak, *Actas del XXXI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, Universidad de Alicante, Alicante, Spain, 2015b. Sociedad Española para el Procesamiento del Lenguaje Natural. ISBN 978-84-608-1989-9.
 - Alkorta J., Gojenola J. eta Iruskieta M. Sentimenduen analisia euskaraz: lexiko-mailatik erlaziozko diskurtso-egiturarako proposamena. *Gogoia: Euskal Herriko Unibersitateko hizkuntza, ezagutza, komunikazio eta ekintzari buruzko aldizkaria (Xabier Arrazola Gogoian (1962-2015))*:14, 131-152, 2016b, ISSN 1577-9424
 - Alkorta J. El uso de información del discurso en el análisis de sentimientos en euskera. In Almela A., Alcaraz-Mármol G., Valencia R., Fernández G. T. *Proceedings of Doctoral Symposium of the 33rd Conference of the Spanish Society for Natural Language Processing (SEPLN 2017)*, Murcia, Spain, 2017c. ISSN 1613-0073
 - Alkorta J., Gojenola J. eta Iruskieta M. 2019b. Towards discourse annotation and sentiment analysis of the Basque Opinion Corpus. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, 144-152, Association for Computational Linguistics. ISBN 978-1-948087-98-8
- **Testuinguruko balentzia-aldatzaileak oro har**
 - Alkorta J., Gojenola J. eta Iruskieta M. 2019a. Sentimenduen tratamendu konputazionalerantz: gramatika maila ezberdinetako sentimendu balentzia-aldatzaileen bila. Olatz Arbelaitz, Urtzi Etxeberria, Ainhoa Latatu, Miren Josu Ormaetxebarria (arg.), *III. Ikergazte. Nazioarteko Ikerketa Euskaraz, Giza Zientziak eta Artea* (1. liburukia), 39-46. Udako Euskal Unibertsitatea (UEU). Bilbo. ISBN: 978-84-8438-682-7
 - **Bestelakoak**

- Alkorta J. Euskararako HPSG gramatikaren lehen proposamena. Alegria I., Latatu A., eta Omaetxebarria M.J., editoreak, *Iker-gazte, Nazioarteko ikerketa euskaraz*, 24-31, Bilbao, Spain, 2015a. Udako Euskal Unibertsitatea. ISBN 978-84-8438-540-0.

- **Bidalitakoak**

- Alkorta J., Taboada M., Gojenola K. eta Iruskietta M. Bidalita. The role of hierarchical structure and coherence relations in sentiment analysis: a study in Basque. *Journal of Corpora and Discourse Studies*

1.5 Tesi-lanaren antolakuntza

Tesi hau zazpi kapituluz osatuta dago. Hurrengo lerroetan, kapitulu bakoitzaren edukia laburbilduta azalduko dugu.

1. kapitulua: proiektuaren nondik norakoak.

Lehen kapituluan, lanaren motibazioa, hipotesi orokorrak, helburuak eta tesiarekin lotutako argitalpenak azaldu ditugu.

2. kapitulua: sentimenduen analisiko baliabideak eta testuinguruko balentzia-aldatzaileak.

Bigarren kapitulua bost atalez osatuta dago. Lehenik eta behin, sentimenduen analisiaren arloan dauden atazak aipatu ditugu. Gero, ataza horiek gauzatzeko behar diren baliabideetako bi (iritzi-testuen corpusak eta sentimenduen lexikoiak) azaldu ditugu. Ondoren, testuinguruko balentzia-aldatzaileak aztergai dituzten lanen inguruan jardungo dugu. Laugarren atalean, sentimenduen analisisa eta euskara uztartzen dituzten lanak aipatu ditugu. Azkenik, kapitulu guztia laburbildu dugu.

3. kapitulua: metodologiaren diseinua.

Kapitulu honetan, lehenik eta behin, euskarazko iritzi-testuen corpora eta euskarazko sentimenduen lexikoa sortzeko urratsak azaldu ditugu. Ondoren, berriz, euskarazko balentzia-aldatzaileak identifikatzeko metodologia zehaztu dugu. Fonologian, morfologian, sintaxian eta diskurtsoan egin dena banan-banan aipatu dugu. Gero, dokumentu mailako sentimenduen sailkatzailea

den SO-CAL tresnaren ingelesezko bertsioa euskaratzeko prozesua nolakoa izan den azaldu dugu. Azkenik, kapituluaren laburpena egin dugu.

4. kapitulua: sentimenduen analisirako baliabideak.

Laugarren kapituluan, tesi-lan honetan zehaztutako metodologia jarraitzearen ondorioz garatutako baliabideen eta beren ebaluazioaren berri eman dugu. Aipatu ditugun baliabideen artean, Euskarazko Iritzi Corpusa, *Sentitegi* izeneko sentimenduen lexikoa eta dokumentu mailako euskarazko sentimenduen sailkatzailea daude. Bukatzeko, kapituluaren laburpena egin dugu.

5. kapitulua: balentzia-aldatzaileak.

Bosgarren kapitulu honetan, lorturiko emaitzen berri eman dugu. Baina aurreko kapitulukoan ez bezala, hemengo emaitzak teorikoak dira. Zehazki, maila fonologiko eta morfologikoan, sintaktikoan eta diskurtsokoan aurkitu ditugun testuinguruko balentzia-aldatzaileak azaldu ditugu eta hitzen edota sintagmen orientazio semantikoan edo sentimenduen balentzian zer-nolako eragina duten ere aipatu dugu. Kapituluarekin amaitzeko, laburpena egin dugu.

6. kapitulua: ekarpenak, mugak eta etorkizuneko lanak.

Kapitulu honek tesi-lanaren ondorioak aipatuko ditu lehenik. Hasierako helburuak bete diren aipatu dugu eta ikerketa-galderei erantzuna emango diegu. Lanaren ekarpenak ere aipatu ditugu. Azkenik, etorkizuneko lanen inguruan aritu gara.

Terminologia eta laburdurak.

Eranskin honetan, tesi-txosten honetan aipatutako terminologia eta laburdurak zerrendatu ditugu.

1. PROIEKTUAREN NONDIK NORAKOAK

AURREKARIAK

Sentimenduen analisiko baliabideak eta testuinguruko balentzia-aldatzaileak

Aurrekarien kapitulu hau bost zatitan banatuta dago. Lehenik eta behin, 2.1 atalean, sentimenduen analisiaren ataza nagusiak zeintzuk diren esplikatzeko da. Ondoren, 2.2 atalean, iritzi-testuen corpusak eta sentimenduen lexikoiak nola sortu diren azaltzen da. 2.3 atalean, berriz, testuinguruko balentzia-aldatzaileekin egin diren lanak aipatzen dira. Euskara eta sentimenduen analisia uztartzen dituzten lanak 2.4 atalean zehazten dira. Azkenik, 2.5 atalean, kapitulu osoa laburtuta dago.

Kapitulu honetan, 2.1 eta 2.2 atalak Liuren (2012) lanean eta bertako sailkapenean oinarrituz egin dira. Sentimenduen analisia eta iritzien erauzketa arloen barnean, hainbat ataza eta hurbilpen daude eta egile bakoitzak sailkapen ezberdin bat egiten du. Guk Liuren (2012) sailkapenaren alde egin dugu; izan ere, arloaren ikuspegi orokorrago bat emateaz gain, garrantzi berezia ematen dio lexikoian oinarritutako edo hizkuntzalaritzaren ikuspegitik gertuago dauden hurbilpenei. Sentimenduen analisiaren inguruan egin diren beste sailkapen batzuk honako lanetan aurki daitezke: Pang eta Lee (2008), Vinodhini eta Chandrasekaran (2012), Cambria *et al.* (2017), Mohammad (2016) edo Westerski (2007), besteak beste.

2.1 Sentimenduen analisiaren atazak

Sentimenduen analisisian, jarraian azaltzen diren hiru ataza nagusi bereizi ohi dira eta ataza horietako bakoitza hizkuntza maila batekin (hitz, esaldi eta diskurtso mailekin) lotuta dago (Liu, 2012).

- Entitate eta aspektu mailako ataza: ezaugarri maila (*feature level*, ingelesez) lantzen da. Hau da, iritzia positiboa edo negatiboa den jakiteaz gain, iritziaren jomuga (entitatea eta bere aspektuak) identifikatu nahi da. Beraz, ataza hau hitz mailakoa da. Lan aipagarrien artean Wiebe *et al.*ena (1999) dago.
- Esaldi mailako ataza: subjektibitatearen sailkapena (*subjectivity classification*, ingelesez). Ataza honetan, esaldiak informazio subjektiboa edo objektiboa (hots, informazio faktuala) duen bereiztea da helburua. Ataza hau ere lantzen duen lanetako bat Wiebe *et al.*ena (1999) da.
- Dokumentu mailako ataza: dokumentu mailako sentimenduen sailkapena (*document-level sentiment classification*, ingelesez). Ataza honen helburua dokumentu baten subjektibitateak balorazio positiboa edo negatiboa duen identifikatzea da. Ataza honetan aipagarriak dira Pang *et al.*ren (2002), Turneyren (2002) eta Pang eta Leeren (2008) lanak. Sentimenduen analisisiko hiru ataza nagusi hauen artean, gure lanak dokumentu mailakoarekin du lotura; izan ere, dokumentu mailako sentimenduen sailkatzailea garatzea dugu helburu.
- Sentimenduen analisisiko beste atazak. Ataza hauek sentimenduen analisiaren eta HPko beste atazen arteko nahasketa dira. Atazak mota askotarikoak dira: iritzi-testuen laburpena, konparaziozko iritzien analisisa, iritzi-bilaketa eta -berreskuraketa, iritzi *spamen* detekzioa eta iritzien kalitatearen neurketa.

Jarraian, lau atazak banan-banan esplikatuko ditugu.

2.1.1. Entitate eta aspektu mailako sentimenduen analisisia

Sentimenduen analisisiko atazetako bat entitate eta aspektu mailako sentimenduen analisisia da. Ataza honetan, iritzia positiboa edo negatiboa den jakiteaz gain, iritziaren hartzailea zein den ere jakin nahi da.

Entitate bati buruzko iritzi bat ematean, entitate horren eta aspektu guztien iritziak beti ez datoz bat. Hori dela eta, beharrezkoa da entitatearen aspektu bakoitzak nolako iritzia duen jakitea. Adibidez, film bat entitate bat da eta bere aspektuak izango lirateke pertsonaiak, atrezzoak, soinu-banda eta argumentua, besteak beste. Aspektu horietako batzuk positiboak izan daitezke eta beste batzuk, berriz, negatiboak. Entitate horren eta aspektu horien orientazio semantikoa identifikatzea da ataza honen motibazioa.

Aspektuak hitz edo sintagma mailan agertzen dira. Horregatik, aspektu bakoitzaren orientazio semantikoa lortzeko, hitz edo sintagmaren orientazioa kalkulatu da. Horretarako, beharrezkoa da iritziak erreferentzia egiten dion entitatea bere aspektuetan deskonposatzea.

Aspektuan oinarritutako sentimenduen analisisian sei urrats daude. Urratsak esplikatzen Liuren (2012) (5) adibidea erabiliko dugu.

- (5) Posted by: bigJohn Date: Sept. 15, 2011. [1] I bought a Samsung camera and my friends brought a Canon camera yesterday. [2] In the past week, we both used the cameras a lot. [3] The photos from my Samy are not that great, and the battery life is short too. [4] My friend was very happy with his camera and loves its picture quality. [5] I want a camera that can take good photos. [6] I am going to return it tomorrow.

- 1- Entitateen erauzketa eta kategorizazioa. Entitate guztiak erauzi eta ezaugarrietan oinarrituta multzokatu egiten dira. (5) adibidean, *Samsung*, *Samy* eta *Canon* entitateak erauzten dira eta *Samsung* eta *Samy* entitateak multzokatzen dira entitate bera errepresentatzen dutelako.
- 2- Aspektuen erauzketa eta kategorizazioa. Aspektu guztiak erauzi eta

multzokatu egiten dira. (5) adibidean, *picture*, *photo* eta *battery life* aspektuak erazten dira eta *picture* eta *photo* aspektuak multzokatzen dira automatikoki, kamerari dagokionez sinonimoak direlako.

- 3- Iritziaren jabearen erazketa eta kategorizazioa. Iritzi-hartzaile guztiak erazi eta beren ezaugarrietan oinarrituta, kategorizatu egiten dira. (5) adibidean, iritziaren jabea [3] esaldian *bigJohn* (blogaren egilea) da eta [4] esaldian, aldiz, iritziaren jabea *bigJohnen* laguna da.
- 4- Denboraren erazketa eta kategorizazioa. Denbora-adierazpenak erazi, formatu bakar batean jarri eta sailkatu egin behar dira. Aztertzen ari garen adibidean, erazitako denbora-adierazpena Sept-15-2011 da.
- 5- Aspektuan oinarritutako sentimenduen analisia. Urrats honetan egiten da sentimenduen analisia, hots, iritziaren aspektuak orientazio positibo edo negatiboa duen zehazten da eta iritziaren hartzaileari kalifikazio bat esleitzen zaio.

[3] esaldian, Samsung kameraren irudi-kalitateari eta bateria-bizitzari iritzi negatiboa ematen zaio. [4] esaldian, iritzi positiboa ematen zaio bai Canon kamerari, bai haren irudi-kalitateari. Gainera, esaldi honetan, *its* eta *this camera* hitzek nori egiten dioten erreferentzia ere jakin behar da. Azkenik, [5] esaldian, iritziak positiboa dela ematen du, baina ez da horrela. Esaldi horretan, hiztunak bere desio edo nahi bat adierazten baitu eta ez objektu baten (kasu honetan, kamera baten) iritzia edo ebaluazioa.

- 6- Aspektu bakoitzari lotutako iritziaren sorkuntza. Azken urratsean, aspektuak, aspektuari loturiko entitatea, aspektuari buruzko iritzia, iritziaren jabea eta iritzia eman den denbora zehazten dira.

(Samsung, *picture_quality*, negative, *bigJohn*, Sept-15-2011)

(Samsung, *battery_life*, negative, *bigJohn*, Sept-15-2011)

(Canon, GENERAL, positive, *bigJohn's_friend*, Sept-15-2011)

(Canon, *picture_quality*, positive, *bigJohn's_friend*, Sept-15-2011)

Hu eta Liuren (2004) lana erreferentziazkoa ataza honetan. Izan ere, bezero askoren iritzi-testuak laburtzeaz gain, laburpen horietatik balorazioren

batzuk dauden zatiak bakarrik hartzen dituzte. Horrela, iritzi datu-base bat hartu eta, lehenik eta behin, POS etiketatzea, ezaugarrien maiztasunaren identifikazioa eta ezaugarrien kimatzea burutzen dute. Ondoren, ohiko ezaugarrietatik iritxidun hitzak lortzen dituzte eta azken honetatik, ezohiko ezaugarriak ere lortzen dituzte. Gero, iritxidun hitzak eta ezohiko ezaugarriak baliatuz, iritxidun esaldiaren orientazio identifikazioa egiten dute eta azkenik, berriz, iritzi-testuaren laburpen bat lortzen dute.

2.1.2. Esaldi mailako sentimenduen analisisa

Sentimenduen analisisiko ataza honek esaldi mailako sentimenduen sailkapena egiten du. Ataza honek bi urrats ditu: i) subjektibitatearen sailkapena eta ii) sentimenduen sailkapena.

Ataza lantzeko azter dezagun Liuren (2012) adibidea hau:

- (6) “[1] I bought a Motorola phone two weeks ago. [2] Everything was good initially. [3] The voice was clear and the battery life was long, although it is a bit bulky. [4] Then, it stopped working yesterday.”

Subjektibitatearen sailkapena izeneko atazean, esaldiak iritzia adierazten duen edo ez zehazten da. Esaldiak iritzia ez du adieraziko baldin eta bertan azaltzen den informazioa faktuala edo objektiboa baldin bada (hau da, informazioa egiazkoa bada). Esaldi batean iritzia, ebaluazioa, emozioa, usteak, espekulazioa, epaia, salaketa edota jarrera, besteak beste, agertzen bada, al-diz, esaldia subjektiboa da. Hala ere, batzuetan posible da esaldi objektibo batek ere iritzia adieraztea¹.

(6) adibidea aztertzen badugu, [1] esaldiak informazio faktuala du; beraz, esaldia objektiboa da. [2] eta [3] esaldiek, berriz, iritziak adierazten dituzte *good*, *clear*, *long* eta *bulky* adjektiboak agertzen baitira. Azkenik, [4] esaldia iritzia adierazten duen esaldi objektiboa da; izan ere, bateriak funtzionatzeari uztea informazio faktuala da baina zerbait negatiboa da.

¹Esaterako, *mugikor honek bateria azkar xahutzen du* esaldia objektiboa da baina iritzi negatiboa ere badu.

Lehen urratsean, subjektibitatearen sailkapena egiteko hainbat teknika aipatu ditugu. Teknika horiek 2.1 taulan laburbilduta eta, ondoren, azalduta daude.

Lana	Teknika
(Wiebe <i>et al.</i> , 1999)	Ikasketa gainbegiratu
(Wiebe, 2000)	Ikasketa gaibegiratu gabea
(Yu eta Hatzivassiloglou, 2003)	Esaldi antzekotasuna eta Naïve Bayes
(Riloff <i>et al.</i> , 2006)	Unigrama, n-grama eta patroï lexiko-sintaktikoen arteko harremanak

2.1 taula: Esaldien subjektibitatearen sailkapena egiteko teknika ezberdinak.

Wiebe *et al.*ek (1999) subjektibitatearen sailkapena egiteko urre-patroia nola sortzen duten azaltzen dute. Etiketatzean desadostasunak agertu zaizkienez, aurreiritziak zuzentzeko etiketak sortzen dituzte bi helbururekin: i) eskuz etiketatzeko gidalerroen bertsio hobea sortzeko eta ii) sailkatzaile automatikoa sortzeko.

Urtebete geroago, Wiebek (2000) subjektibitatearen sailkapena egiteko ikasketa gainbegiratu gabea² egiten du. Kasu honetan, hitz-multzoen distribuzio semantikoa aintzat hartuz, subjektibitateari loturiko aztarnak lortzen ditu.

Yu eta Hatzivassiloglouk (2003), berriz, iritzidun esaldiak bilatzeko hiru teknika ezberdin erabiltzen dituzte: i) esaldien arteko antzekotasuna, ii) Naïve Bayes sailkatzailea³ eta iii) Naïve Bayes anizkuna.

Bukatzeko, Riloff *et al.*ek (2006) ezaugarrien errepresentazioa egiteko subsumzio hierarkia (*subsumption hierarchy*) erabiltzen dute, bertan, ezaugarri

²Ikasketa gainbegiratu gabea ikasketa automatikoko metodo bat da, behaketan oinarritzen den eredu bat da. Bertan, modelatze prozesu guztia sistemako sarreran dauden adibide multzo baten bidez gauzatzen da. Ez du adibide horien kategorien inguruko informaziorik. Hori horrela izanik, sistemak gauza izan behar du patroïak identifikatu eta, modu horretan, sarrera berriei etiketak esleitzeko.

³Naïve Bayes sailkatzaile probabilitistiko bat da eta Bayes-en teoreman eta aldagaien arteko independentziaren hipotesian oinarritzen da. Sailkatzaile honen arabera, klase aldagai bat emanda, ezaugarriak elkarren artean independenteak dira eta, beraz, ez dago korrelaziorik.

lexikal ezberdinak eta beren arteko harremanak finkatzeko. Modu horretan, esaldien subjektibitatearen sailkapena lortzen dute.

Bigarren urratsean, berriz, sentimenduen sailkapena egiten da eta esaldiak iritzia adierazten badu, iritzia positiboa edo negatiboa den zehatzen da.

(6) adibidearen kasuan, [2] esaldiak iritzi positiboa du, *good* (“ona”) hitza dagoelako, baina [3] eta [4] esaldiek iritzi negatiboa dute: [3] esaldiaren kasuan, *bulky* (“tamaina handikoa”) hitza ageri da eta [4] esaldian, azkenik, mugikorrek funtzionatzeari utzi diola aipatzen da.

Sentimenduen sailkapena egiteko ere hainbat teknika daude eta horiek 2.2 Taulan laburbilduta eta, ondoren, azalduta daude.

Lana	Teknika
(Yu eta Hatzivassiloglou, 2003)	Egiantz-arrazoiaren logaritmoa (<i>log-likelihood ratio</i> , LLR)
(Hu eta Liu, 2004)	Lexikoian oinarritutako algoritmoa
(Gamon, 2004)	Ikasketa algoritmo erdi-gainbegiratua
(McDonald <i>et al.</i> , 2007)	Sekuentzia hierarkikoen ikasketa automatikoa

2.2 taula: Esaldi mailako sentimenduen sailkapena egiteko zenbait teknika.

Sentimenduen sailkapenaren urratsa gauzatzeko, Yu eta Hatzivassiloglouek (2003) hitzen orientazio semantikoa kalkulatzeko eta horretarako Turneyk (2002) proposaturiko teknika erabiltzen dute. Hurrengo urratsean, esaldi osoaren orientazio semantikoa kalkulatzeko, berriz, egiantz-arrazoiaren logaritmoa⁴ erabiltzen dute.

Hu eta Liuek (2004), aldiz, hiru urrats egiten dituzte sentimenduen sailka-

⁴Egiantz-arrazoiaren logaritmoak (*log-likelihood ratio*, ingelesez) bi eredu estatistikoaren egokitzearen egokitasuna ebaluatzen du. Bi ereduok beren probabilitateen proportzioaren arabera lehiatzen dira; zehazki, bata parametroen espazio osoaren maximizazioarekin aurkitzen da eta bestea murriztapen batzuk ezarri ondoren aurkitzen da. Behatutako datuek murriztapena babesten badute (hau da, hipotesi nulua) bi probabilitateek ez dute laginketa-errore bat baino gehiagotan diferitu behar. Beraz, egiantz-arrazoiak erlazio hau bata bestearengandik oso ezberdina den edo bere logaritmo naturala zerotik guztiz ezberdina den ziurtatzen du.

pena gauzatzeko: i) kontsumitzaileen iruzkinetan agertu diren produktuaren ezaugarriak erauzi, ii) iruzkin bakoitzean iritzidun esaldia identifikatu eta positiboa edo negatiboa den adierazi eta iii) emaitzak laburbildu.

Bestalde, Gamonek (2004) ikasketa erdi-gainbegiratuen teknika erabiltzen du. SMV algoritmoa ezaugarri askoko bektoreekin erabiltzen du, zeintzuek bektore murrizketa jasan duten. Halaber, azterketa linguistiko sakonaren ezaugarriak hitzen n-grama funtzioei ere aplikatzen dizkie.

Azkenik, McDonald *et al.*ek (2007) testuaren granularitate-maila ezberdinetan testuen sailkapen semantikoa egiteko eredu egituratuak erabiltzen dituzte. Ereduaren inferentzia Viterbiren sekuentzia estandarren sailkapen teknikan oinarritzen dira.

Ikusten den moduan, teknika ezberdinak daude esaldi mailako sentimenduen analisia egiteko, hots, esaldien iritzi positiboa edo negatiboa duten jakiteko.

2.1.3. Dokumentu mailako sentimenduen sailkapena

Aurretik esan bezala, ataza honetan dokumentu baten eduki subjektiboa aztertzen da, duen balorazioa positiboa edo negatiboa den identifikatuz. Ataza honetarako hainbat hurbilpen daude (Liu, 2012): i) sentimenduen sailkapena ikasketa gainbegiratu erabiliz (2.1.3.1 atala), ii) sentimenduen sailkapena ikasketa gainbegiratu gabea erabiliz (2.1.3.2 atala), iii) sentimenduen estimazio iragarpena (2.1.3.3 atala), iv) domeinu arteko sentimenduen sailkapena eta (2.1.3.4 atala) v) hizkuntza arteko sentimenduen sailkapena (2.1.3.5 atala).

2.1.3.1. Sentimenduen sailkapen gainbegiratu

Sentimenduen sailkapena ikasketa gainbegiratu egitean, dokumentu baten sentimendu sailkapen bitarra egiten da, hots, haren iritzia positiboa edo negatiboa den zehazten da. Zehaztapen hori kalifikazio bat emanez egiten da. Ikasketa gainbegiratuan, entrenamenduko eta testeko datuak iritzi-testuak dira eta bertan, iritzia 1 eta 5 bitarteko eskala erabiliz adierazten da non 1 eta 2 kalifikazioak negatiboak diren eta 4 eta 5 kalifikazioak, aldiz, positiboak. Ikertzaile gehienek ez dute iritzi neutroa aintzat hartzen, modu horretan,

sailkapena egitea erraza delako, baina beharko balitz, 3 kalifikazioa izango litzateke iritzi neutroa.

Sentimenduen sailkapena bereziki dokumentu mailako arazo bat da. Testu-sailkapen tradizionalan, hots, testu bat domeinu batean sailkatzeko arazoan, domeinu ezberdinetako dokumentuak sailkatu izan dira: politika, zientzia eta kirola, besteak beste, eta mota horretako sailkapenetan, domeinuari loturiko hitz gakoak garrantzitsuak dira, modu horretan, testua gai nagusiari edo bigarren mailako gai bati buruzko iritzia ematen ari den jakin daitekeelako. Baina sentimenduen sailkapenean, iritzi positiboa edo negatiboa adierazten duten sentimendu eta iritzidun hitzak garrantzitsuagoak dira (esaterako, *ederra*, *bikaina*, *txarra*), Liuk (2012)

Dokumentu mailako sentimenduen sailkapena testu-sailkapenaren ataza bat denez, ikasketa gainbegiratuan dauden teknika ezberdinak erabil daitezke. Dokumentu mailako sentimenduen sailkapena egiteko ikasketa automatiko gainbegiratuak aplikazioek zenbait ezaugarri erabili izan dituzte:

- 1- Hitzak eta horien maiztasunak. Banakako hitzek (unigramak) eta beren n-gramak⁵ maiztasun-kontaktarekin osatzen dute ezaugarri hau. Batzuetan, posizioa ere kontutan hartzen da. Ezaugarri honek eragin-kortasun handia erakusten du sentimenduen analisisian.
- 2- *Part-Of-Speech* (POS). Hitz bakoitzaren gramatika-kategoria jakitea ere garrantzitsua da; izan ere, sentimenduen analisisian gramatika-kategoria bakoitzak garrantzi ezberdina du. Hala, adjektiboak iritziaren adierazle garrantzitsuak dira. Adjektiboek halako garrantzia dutenez, batzuetan ezaugarri berezi moduan tratatzen dira. Beste batzuetan, aldiz, n-gramak eta POS batera erabiltzen dira ezaugarri berean.
- 3- Sentimendu-hitzak eta -sintagmak. Sentimendu hitzak iritzi positiboak eta negatiboak adierazteko erabiltzen dira. Adibidez, *on*, *zorioneko* eta *eder* hitz positiboak dira. *txar*, *zoritxarreko* eta *itsusi*, aldiz, hitz negatiboak dira. Sentimendu-hitz gehienak adjektiboak eta adberbioak

⁵N-grama sekuentzia baten parte den azpi-sekuentzia bat da. Azpi-sekuentzia hori hainbat n elementuz osatuta dago eta elementu horiek hainbat motatakoak izan daitezke; letrak edo hitzak, esaterako.

dira baina izenak (*maitasuna*) edota aditzak (*gorrotatu*) ere izan daitezke. Sentimendu-hitzez gain, sentimendu-sintagmak ere badaude; *leher eginda* lokuzioa oso nekatuta zaudela adierazteko bezalakoak.

- 4- Iritzi-erregelak. Sentimendu-hitz eta -sintagmez gain, badaude iritziak adierazteko zenbait adierazmolde iritzia edo sentimendua adieraz edo inplika dezaketenak. Esaterako, *bateria-bizitza luzea du* eta *bateria-bizitza laburra du* esaldietan, aurkako bi orientazio semantiko daude; baina halaber, *bateria-bizitza* hitzak berak orientazio semantiko positiboa inplikatzeko du. Hau da, mugikorrek *bateria-bizitza* bera edukitzea positiboa da baina horren baitan, *bateria-bizitza luzea* positiboa den bezala *bateria-bizitza laburra* negatiboa da.
- 5- Sentimendu-aldatzaileak. Badaude zenbait adierazmolde orientazio semantikoa aldatzeko erabiltzen direnak. Orientazio semantiko positiboa negatibo bilaka dezakete edota alderantziz. Sentimenduen aldatzaileetako bat ezeztapena da. *Kamera hau ez zait gustatzen* esaldian, esaterako, *gustatu* hitzaren orientazio semantiko positiboa negatibo bihurtzen du.
- 6- Dependentsia sintaktikoak. Dependentsia-zuhaitzetatik sortutako hitzen dependentsia-ezaugarriak ere erabiltzen dira dokumentu mailako sentimendu-sailkapenean.

Dokumentu mailako sentimenduen sailkapena egiteko erabilitako teknikak 2.3 Taulan daude laburbilduta.

Aurretik aipatutako ezaugarriak erabiliz, zenbait teknika erabili izan dira. Pang *et al.*ek (2002) filmeen iruzkinak testu positibo eta negatibo bezala sailkatzen dituzte eta horretarako n-gramak (hitz-poltsak) erabiltzen dituzte sailkapeneko ezaugarri moduan Naïve Bayes nahiz SVM sailkatzaileak erabiliz.

Mullen eta Collierek (2004) SVM sailkatzailea⁶ hainbat ezaugarriekin konbi-

⁶Euskarri bektoredun makina (ingelesez, *Support Vector Machine*, SVM) sailkapen nahiz erregresiorako erabiltzen den algoritmo multzoa da. Bere oinarrian bektore-

Lana	Teknika
Pang <i>et al.</i> (2002)	1) Naïve Bayes sailkatzailea 2) Euskarri bektoredun makina (<i>Support Vector Machines</i> , SVM)
Mullen eta Collier (2004)	SVM + Erlazio sintaktikoak eta ohiko ezaugarriak
Nakagawa <i>et al.</i> (2010)	Dependentzia-zuhaitzean oinarritutako sailkatzailea + baldintzazko ausazko eremuak (<i>Conditional Random Fields</i>), CRF
Kouloumpis <i>et al.</i> (2011)	Ezaugarri linguistikoak + mikroblogetako hizkuntza kreatiboa eta ez-formala identifikatzeko ezaugarria

2.3 taula: Dokumentu mailako sentimenduen sailkapena egiteko erabiltako zenbait teknika.

natzen dute, hala nola PMI⁷ bidez lorturiko orientazio semantikoak, Wordneten oinarrituz egindako desberdintze semantikoa eta gaiarekiko hurbiltasuna eta erlazio sintaktikoen ezaugarriak.

Nakagawa *et al.*ek (2010), berriz, ingeleseko eta japonierako datu subjektiboak sailkatzeko dependentzia-zuhaitzak eta baldintzazko ausazko eremua⁸ erabiltzen dituzte. Lanean, esaldi bakoitzaren azpi-zuhaitz baten orientazio

espazioak daude eta sarreran, bere n dimentsio eremuan adieraz daitezkeen hainbat datu jasotzen ditu, bi kategorietatik batean sailkatuta daudenak. Ondoren, teknika honek bere eremuan dauden puntuak bereizteko hiperplanoa bilatzen du. Teknika honen kasuan gertueneko puntuetarako distantziarik handiena duena aukeratzen du eta hori da emaitza.

⁷Elkarrekiko Informazio Puntuala (ingelesez, *Pointwise mutual information*) estatistikan eta informazioaren teorian erabiltzen den asoziazio neurria da. Hitz batek (x) beste hitz batekin (y) duen probabilitatea hitzaren beraren (x) probabilitatearekin zatituz lor-tzen da. Esaterako, *puerto* hitza 1.938 aldiz agertzen da eta *rico* 1.311 aldiz. Biak batera, berriz, 1.159 aldiz agertzen dira. Horrenbestez, PMI neurria 10,03 da.

⁸Baldintzazko ausazko eremuak (ingelesez, *Conditional Random Fields*, CRF) dokumentuetatik datuak erauzteko edo datu-sekuentziak segmentatu eta etiketatzeko erabiltzen den eredu estokastikoa da. Eredu honek datu-sekuentzia bat emanda elementuetako bakoitzari (O_i) etiketa bat (S_i) ematen dio. Eredu sortzailea denez, behaketen eta etiketen probabilitatearen banaketa aldi berean modelatzen ditu eta baldintzazko ausazko eremuek behaketek baldintzatutako etiketen sequentzia zuzenen probabilitatea modelatzen dute ($P(S-O)$). Beraz, eredu diskriminatzaile bat da.

semantikoa aldagai bat da eta esaldi osoaren orientazio semantikoa aldagai horien arteko interakzioa kalkulaturaz lortzen da.

Kouloumpis *et al.*ek (2011), berriz, txioen sentimenduen sailkapena egiteko, hainbat ezaugarri erabiltzen dituzte: n-gramen ezaugarriak, lexikoiko ezaugarriak (hitzak positiboak, negatiboak edo neutralak diren ere barne hartuz), gramatika-kategoria ezaugarriak eta, bukatzeko, mikroblogetako testuen ezaugarriak.

2.1.3.2. Sentimenduen sailkapen gainbegiratu gabea

Dokumentu mailako sentimenduen sailkapena ikasketa gainbegiratu gabea erabiliz egiten duten lanak ere badaude. Sentimendu-hitzak oinarritzko faktoreak dira sentimenduen sailkapenean eta horregatik, horiek modu gainbegiratu gabean erabil daitezke. Bi hurbilpen daude ikasketa gainbegiratu gabeen: i) erregela sintaktikoetan oinarritutakoa eta ii) lexikoian oinarritutakoa.

Bi hurbilpenak argitzeko Liuren (2012) *this piano produces beautiful sounds* esaldia erabiliko dugu.

	Lehen hitza	Bigarren hitza	Hirugarren hitza (ez erauzia)
1	JJ	NN edo NNS	edozer
2	RB, RBR edo RBS	JJ	ez NN edo NNS
3	JJ	JJ	ez NN edo NNS
4	NN edo NNS	JJ	ez NN edo NNS
5	RB, RBR edo RBS	VB, VBD, VBN edo VBG	edozer

2.4 taula: Dokumentu mailako sentimenduen sailkapen gainbegiratu gaberako erregela sintaktikoak (Turney, 2002).

Turneyk (2002) erregela sintaktikoen hurbilpena jarraitzen du. Erregela sintaktikoak gramatika-kategorien POS etiketetan (Liu, 2012)⁹ oinarritzen dira

⁹Hauexek dira erregeletan azaltzen diren laburpenen esanahiak: JJ: adjektiboa, RB: adberbioa, RBR: konparaziozko adberbioa, RBS: superlatibozko adberbioa, NN: izen singularra, NNS: izen plurala, VB: aditza erro moduan, VBD: aditza iraganaldian, VBN: aditza iragan partizipioan eta VBG: aditza orainaldian eta orainaldiko partizipioan.

eta guztira hiru urrats daude hurbilpen honetan:

1. urratsa. Jarraian dauden bi hitz erauzi, baldin eta bi hitz horien POS etiketek 2.4 taulako patroia bat osatzen badute. Esaterako, bigarren patroiak adierazten du lehen hitzak adberbioa, bigarrenak adjektiboa eta hirugarrenak (erauzten ez denak) izena izan behar duela. *This piano produces beautiful sounds* esaldiaren kasuan, *beautiful sounds* dugu, lehen hitza adjektiboa da eta bigarrena, berriz, izena. Beraz, 2.4 taulako lehen patroiarekin bat egiten du eta bi hitzak erauzi egiten dira. JJ, RB, RBR eta RBS POS etiketak dituzten hitzek askotan iritzia adieraztea da patroiak erabiltzearen arrazoia eta berekin agertzen diren izen eta aditzek beren testuingurua dira. Izan ere, adjektibok esaterako (*ikaragarri* kasu) testuinguruaren arabera sentimendu positiboa edo negatiboa izan dezakete.
2. urratsa. Aurretik erauzitako hitz-multzoaren orientazio semantikoa kalkulatzeko da elkarrekiko informazio puntuala (PMI) neurria (2.1 irudia) erabiliz.

$$PMI(term_1, term_2) = \log_2 \left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1)\Pr(term_2)} \right)$$

2.1 irudia: *Pointwise mutual information*, PMI.

2.1 irudiko PMI neurriak bi hitzen arteko dependentzia estatistikoa neurtzen du. $\Pr(term_1 \wedge term_2)$ *term1* gertatzeko ko-okurrentzia probabilitatea eta $\Pr(term_1)\Pr(term_2)$ bi hitzen ko-okurrentzia probabilitatea da, baldin eta bi hitzak estatistikoki independenteak badira.

$$(7) \quad SO(\text{phrase}) = PMI(\text{phrase}, \text{“beautiful”}) - PMI(\text{phrase}, \text{“ugly”}).$$

Liuren (2012) *this piano produces beautiful sounds* sintagmaren orientazio semantikoa (SO) kalkulatzeko, (7) adibidean ikusten den moduan, erreferentzia positiboko hitza den *beautiful* eta negatibokoa den *ugly* hitzaren arteko asoziazioa neurtzen da.

Probabilitatea kalkulatzeko, bilaketa-motor¹⁰ bati kontsulta eskatzen zaio eta bisita kopurua biltzen du hark. Bilaketa bakoitzeko kontsulta batean, bilaketa-motor batek kontsultari lotutako dokumentu kopuru aipagarriak eskaintzen ditu eta hori emaitza kopurua da. Beraz, bi hitz batera eta bereiz bilatuta 2.1 irudiko probabilitatea kalkula daiteke.

3. urratsa. Iritzi-testu bat emanda, algoritmoak iritzi-testuan dauden sintagma guztien orientazio semantikoa neurtzen du eta iritzi-testua positibotzat (orientazio semantiko guztien batezbestekoa positiboa bada) edo negatibotzat (orientazio semantiko guztien batezbestekoa negatiboa bada) sailkatzen du.

Taboada *et al.*ek (2011) ere ikasketa gainbegiratu gabea erabiltzen du. Bertan, dokumentu mailako sentimenduen sailkapena egiteko SO-CAL tresna aurkezten du¹¹. Tresna honek hitz, esaldien eta testuen orientazio semantikoa kalkulatzeko du.

Tresna honen abiapuntua -5 eta $+5$ arteko orientazio semantikoa duten hitzen lexikoa da. Guztira, lexikoiak lau gramatika-kategorietako hitzak biltzen ditu: izenak, adjektiboak, adberbioak eta aditzak. SO-CAL tresnak hitz horiek iritzi-testu batean bilatzean erauzi egiten ditu. Hurrengo urratsean, tresnak hizkuntzaren beste ezaugarri batzuk erauzten ditu, hala nola intentsifikatzaileak, ezeztapena edota puntuazio-zeinuak. Erauzketa egin ostean, lehen urratsean erauzitako hitzen sentimendu-balentziak aldaketa batzuk jasaten ditu, aipatu berri ditugun hizkuntza fenomenoen ondorio. Aldaketa hori sentimendu-balentzia indartzea edo ahultzea izan daiteke.

Esaterako, intentsifikatzaileek hitzaren orientazio semantikoa indartu edo ahuldu dezake. Bestalde, sintagman edo esaldian ezeztapen-marka agertuz gero, horren orientazio semantikoa ahuldu egiten du tresnak, ezeztapen-markaren ± 4 balentzia aplikatuz. Hau da, sintagma edo esaldiaren orienta-

¹⁰Bilaketa-motorra informazioa bilatzeko helburuarekin sortutako informazioa eskuratzeko sistema da. Baldintza batzuk zehazten dira motorrean eta horren arabera emaitzak lortzen dira zerrenda batean garrantzi handienetik txikienera.

¹¹Sentimenduen sailkapena egiten duen SO-CAL tresnari buruzko informazio gehiago: atal hauetan dago: 4.3.1 atalean non tresnari buruzko informazio gehiago ematen den eta 4.3.2 atalean non funtzionamendu bera duen euskarazko bertsioa deskribatzen den.

zio semantikoa negatiboa bada eta ezeztapena badago, ezeztapen-markaren +4 aplikatuz, haren orientazio semantikoa indartu egingo du. Sintagma edo esaldiaren orientazio semantikoa positiboa bada eta ezeztapena badago, aldiz, ezeztapen-markaren -4 balentziak, haren orientazio semantikoa ahulduko du.

Bukatzeko, SO-CAL tresnak puntuazio-zeinuak ere kontuan hartzen ditu eta, esaterako, komen artean dauden hitzei ez die orientazio semantikoa esleitzen, izenburuak edota aipuak bezalakoak direlako.

Liuren (2012) esaldia hurbilpen honetan aplikatzen badugu, lehenik eta behin, SO-CALen lexioiak esaldiko *beautiful* (+4) hitza erauzten du. Hurrengo urrartsean, inguruan intentsifikatzailerik eta ezeztapenik ez dagoela antzematen du, eta horrenbestez, hitzaren orientazio semantikoari ez dizkio hizkuntza-fenomeno hauen eraginak aplikatzen. Bukatzeko, tresnak hitza komen artean ez dagoela egiaztatzen du eta horrenbestez, hitzaren orientazio semantikoa bere horretan mantentzen du. Esaldian, orientazio semantikodun beste hitzik ez dagoenez, *beautiful* hitzaren orientazio semantikoa da (+4) esaldiaren orientazio semantikoa.

Gure tesi-lanak sentimenduen analisi gainbegiratu gabearekin du zerikusia. Izan ere, garatu nahi dugun sentimendu-sailkatzailea lexikoian oinarritutakoa izango da.

2.1.3.3. Sentimenduen estimazio-iragarpena

Dokumentu mailako sentimenduen sailkapeneko beste arloetako bat sentimenduen estimazio-iragarpena (*Sentiment Rating Prediction*) da. Arlo honetan, dokumentu batek iritzi positiboa edo negatiboa duen sailkatzetik urrunago joan nahi da eta zenbakizko kalifikazio bat (1 eta 5 artean) esleitu nahi zaio. Kasu honetan, ataza erregresio¹² moduan trata daiteke; izan ere, kalifikazio-puntuazioak ordinalak dira.

Sentimenduen estimazio iragarpena-egiteko aipatuko ditugun teknika ezberdinak 2.5 taulan daude laburbilduta.

Pang eta Leek (2005) SVM erregresioa erabiltzen du. Zehazki, klase anitzeko

¹²Erregresioa aldagai dependente baten eta independente zenbaiten arteko erlazioa zehaztu eta aztertzen duten teknika estatistikoa da.

Teknika	Lana
SVM erregresioa	Pang eta Lee (2005)
Hitz-poltsen errepresentazioa	Qu <i>et al.</i> (2010)
Aspektuaren sentimenduen estimazio-iragarpena: Erregresio estandarra	Snyder eta Barzilay (2007)
Aspektuaren sentimenduen estimazio-iragarpena: Sare Bayesiarrean oinarritutako sailkatzailea	Long <i>et al.</i> (2010)

2.5 taula: Sentimenduen estimazio-iragarpena egiteko teknika ezberdinak.

SVM sailkatzailea¹³ erabiltzen du.

Bestalde, Qu *et al.*ek (2010) dokumentuko hitz-poltsen errepresentazioa darabilte iritzia duten n-gramen indarra neurtzeko. Berau ezberdina da ohiko hitz-poltsen errepresentaziotik. Teknika honetan iritzietako bakoitza hirukoitza da: sentimendu hitza, aldatzailea eta ezeztapena baititu. Esaterako, *ez oso ona* adibidean, *ez* ezeztapena da, *oso* aldatzailea eta *ona*, berriz, sentimendu-hitza. Modu honetan, sentimendu-hitzetik haratago dauden baino berengan eragiten duten fenomenoak identifika ditzakete.

Snyder eta Barzilayk (2007) beste ikuspegi batekin lantzen dute ataza. Iritzi bakoitzari zenbakizko kalifikazio bat esleitu beharrean, aspektuetako bakoitzari esleitzen diote kalifikazioa eta, horretarako, erregresio estandarra edo sailkatze-teknikak erabiltzen dituzte.

Azkenik, Long *et al.*ek (2010) iritzietako aspektu bakoitzari kalifikazioa esleitzeko sare Bayesiarra darabilen sailkatzailea erabiltzen dute.

2.1.3.4. Domeinuen arteko sentimenduen analisisia

Dokumentu mailako beste arlo bat domeinu arteko sentimenduen sailkapena da. Ikasketan erabilitako domeinuek eragin handia dute sentimenduen sailkatzaileetan. Horregatik, domeinu jakin bateko iritzi-testuekin ondo dabilen sentimenduen sailkatzaile bat, beste domeinu bateko iritzi-testuekin ez da

¹³Klase anitzeko SVM teknikan, datuak bi kategoria baino gehiagotan sailkatzeko bi hurbilpen daude: i) kategoria bakoitza beste batzuetan zatitzen da eta denak konbinatzen dira eta ii) $k(k-1)/2$ ereduak sortzen dira non k kategoria kopurua den.

ondo ibiltzen (Liu, 2012).

Hitz batek aurkako bi orientazio semantiko eduki ditzake bi domeinu ezberdinetan eta hori da arazo honen jatorria. Adibidez, *luzea* adjektiboaren kasuan, telefono mugikorren domeinuan bateria-bizitzari zuzenduta dagoenean orientazio positiboa du. Adjektibo berak, aldiz, zinemaren domeinuan orientazio negatiboa du: *filma luzea egin zitzaidan*.

Arlo honetan bi hurbilpen daude. Horietako bat iturri-domeinua (*source domain*) da. Gamon *et al.* (2005) da lan aipagarrietako bat hurbilpen honetan, etiketatutako datu ikasiak domeinu berri batean erabiltzen dira. Beste hurbilpena helburu-domeinua (*target domain*) da. Blitzer *et al.* (2007); Tan *et al.* (2005) dira lanik nabarmenenak eta kasu honetan, domeinu berri baterako ez dute etiketatutako daturik erabiltzen.

2.1.3.5. Hizkuntza arteko sentimenduen analisisia

Orain arte deskribatu eta aipatu ditugun hurbilpenak eta lanak hizkuntza bati bideratuta egon dira. Hala ere, badira hizkuntza arteko sentimenduen analisisia egiten duten hurbilpen eta teknikak ere, bai esaldi mailan, bai dokumentu mailan.

Esaldi mailako hizkuntza arteko sentimenduen analisisian ere bi ataza daude: i) hizkuntza arteko subjektibitatea eta ii) sentimenduen sailkapena.

Arlo honetan, hizkuntza askotan baliabide falta dagoenez, ingelesetik beste hizkuntzetara sentimenduen analisisirako baliabide bat automatikoki itzultzea da helburua. Liu (2012) lanaren arabera, hiru estrategia nagusi daude baliabideak itzultzerakoan:

- 1- Helburu-hizkuntzako test-multzoko esaldiak iturri-hizkuntzara itzuli eta bertan, iturri-hizkuntza iturriko hizkuntzan dagoen baliabidea edo sistema erabiliz test-multzoko esaldiek informazio objektiboa edo subjektiboa duten sailkatu.
- 2- Iturri-hizkuntzako ikasketa corpora¹⁴ helburu-hizkuntzara itzuli eta corpusen oinarritutako sailkatzailea sortu helburu-hizkuntzan.

¹⁴Ikasketa automatikoan, corpora hiru zatitan banatu ohi da: garapena, ikasketa eta

- 3- Iturri-hizkuntzako sentimenduen lexikoa helburu-hizkuntzara itzuli eta lexikoian oinarritutako sailkatzailea sortu helburu-hizkuntzan.

Hurbilpen honetan aipaturiko lanak 2.6 taulan laburbilduta daude.

Teknika	Lana
Helburu-hizkuntzako datuak jatorri-hizkuntzara itzuli eta sentimenduen sailkapena egin	Kim eta Hovy (2006) Banea <i>et al.</i> (2008)
Sentimenduen sailkapena egiteko lexikoa helburu-hizkuntzara itzuli eta, gero, sentimenduen sailkapena egin	Kim eta Hovy (2006) Banea <i>et al.</i> (2008) Bautin <i>et al.</i> (2008)
Sentimenduen sailkapena jatorri-hizkuntzan eta, ondoren, itzulpena egin	Banea <i>et al.</i> (2008)
Eleaniztasunaren konparagarritasuna	Kim <i>et al.</i> (2010)

2.6 taula: Esaldi mailako hizkuntza arteko sentimenduen analisirako zenbait teknika.

Kim eta Hovyek (2006) bi esperimendu gauzatzen dituzte: i) alemanerazko emailak ingelesera itzuli eta itzulpenei ingelesezko sentimendu-hitzak aplikatu, horien orientazio semantikoa ikusteko eta ii) ingelesezko sentimendu-hitzak alemanerara itzuli eta itzultakoa alemanerazko emailetan aplikatu.

Banea *et al.*ek (2008) hiru esperimendu egin dituzte: i) ingelesezko datu etiketatutak (iturri-hizkuntza) automatikoki errumanierara itzuli (helburu-hizkuntza), ii) jatorri-hizkuntza automatikoki subjektibitatez etiketatu eta ondoren helburu-hizkuntzara itzuli eta iii) helburu-hizkuntza jatorri-hizkuntzara itzuli eta bertan subjektibitatearen sailkapen tresna aplikatu.

Bautin *et al.*ek (2008), berriz, aspektuan oinarritutako hizkuntza arteko sentimenduen analisia egiten dute. Helburu-hizkuntza iturri-hizkuntzara itzultzen dute (ingelesera) eta ondoren, entitatea duen esaldi bakoitzari orientazio semantikoa esleitzen diote ingelesezko lexikoian oinarritzen den metodoa erabiliz.

testa. Ikasketa zatia, hitzak berak adierazten duen moduan, ikasketa automatikoko teknikek esperientzia trebetasun edo jakintza bihurtzeko baliatzen dute.

Bukatzeko, Kim *et al.*ek (2010) eleaniztasunarekin konpagarritasuna (ingelesez, *multilanguage-comparability*) metodoa aplikatzen dute. Hots, testu elea-nitz bateko hitz-bikoteek subjektibitate esanahi bera dutela baliatuz, sailka-pen emaitzen hitz-bikoteek duten adostasun-maila neurtzen dute.

Dokumentu mailako hizkuntza arteko sentimenduen analisia ere badago. Hizkuntza arteko sentimenduen analisia egiteko bi motibazio nagusi daude:

- i) Ikertzaile batek sentimenduen analisirako sistema bere hizkuntzan sor-tu nahi du eta, bere hizkuntzan horrelakorik sortzeko baliabiderik ez dagoenez eta sistema gehienak ingelesez daudenez, ingeleseko sisteman oinarrituz bere hizkuntzan sistema sortzeko erabiltzen du.
- ii) Aplikazio edo produktuetan, enpresek pertsonen iritziak jaso nahi di-tuzte eta, hori lortzeko, ingelesez duten sistema beste hizkuntzetara itzultzen dute.

Dokumentu mailan, aipatuko ditugun hizkuntza arteko sentimenduen anali-sirako teknika ezberdinak 2.7 taulan laburbilduta daude.

Teknika	Lana
Hainbat itzultzaile automatiko	Wan (2008)
Ko-ikasketa metodoa (SVM barne)	Wan (2009)
Transferentzia-ikaskuntza metodoa	Wei eta Pal (2010)
Gai-eredua	Guo <i>et al.</i> (2010)

2.7 taula: Dokumentu mailan, hizkuntza arteko sentimenduen analisia egiteko zenbait teknika.

Wanek (2008) ingeleseko baliabideak ustiatzen ditu txinerazko iritzi-testuen sentimenduen sailkapena egiteko. Bere algoritmoan, txinerazko iritzi-testuak hainbat itzultzaile erabiliz ingelesera itzultzen ditu eta itzultitako testuei in-geleseko sentimenduen lexikoia aplikatzen die.

Wanek (2009), ordea, ko-ikasketa metodoa¹⁵ erabiltzen du. Zehazki esanda,

¹⁵Ikasketa automatikoan, ko-ikasketa metodoa (ingelesez, *co-training*) etiketatutako datu-multzo txiki bat eta etiketatu gabeko datu-multzo handi bat erabiltzean datza.

txinerazko iritzi-testuen sailkapena egiteko etiketatuta dagoen ingelesezko corpora erabiltzen du modu gainbegiratu batean (SVM teknika erabiliz).

Wei eta Palek (2010), berriz, transferentzia-ikaskuntza metodoa¹⁶ proposatzen dute hizkuntza arteko sentimenduen analisia egiteko. Itzulpen-prozesuan bertan, itzultzaile automatikoak sortzen duen zarata txikiagotzen du.

Azkenik, Guo *et al.*ek (2010) gai-ereduan oinarritzen den metodoa¹⁷ proposatzen dute hizkuntza askotako iritzi-adierazpenak multzokatu eta, modu horretan, estatu askotako iritzien aspektuan oinarritutako sentimenduen sailkapena egiteko.

2.1.4. Sentimenduen analisiko beste atazak

Liuk (2012) adierazten duenez, dokumentu maila, esaldi maila eta entitate eta aspektu maila sentimenduen analisiko ataza nagusiak dira; izan ere, maila horietako bakoitza objektiboa edo subjektiboa den zehaztu, horien orientazio semantikoa zehaztu eta azkenik, iritziaren hartzailea eta entitate baten aspektu askoren orientazio semantikoa zehazten baita.

Hala ere, sentimenduen analisia zabalagoa da eta badaude bestelako atazak ere non sentimenduen analisia hizkuntza naturalaren prozesamenduko beste arloekin uztartzen den. Ataza horiek gure tesi-lanaren helburuetatik urrutiago daudenez, azaletik esplikatuko ditugu.

¹⁶Transferentzia-ikaskuntza (ingelesez, *transfer learning*) ikasketa automatikoko ikerketa arazo bat da eta arazo bat ebatzi ostean lorturiko jakintza gorde egiten du; ondoren, arazo ezberdin baina antzeko batean aplikatzeko. Adibidez, autoak sailkatzeko erabilitako jakintza kamioiak sailkatzeko arazoan erabil daiteke. Psikologian dagoen transferentzia-ikaskuntzarekin nolabaiteko harremana du.

¹⁷Ikasketa automatikoan eta lengoia naturalaren prozesamenduan gai-eredua (ingelesez, *topic model*) testu multzo batean ageri diren “gai” abstraktuak ezagutzeko eredu estatistiko mota bat da. Testuen baitan dauden ezkutuko egitura semantikoak ezagutzeko maiz erabiltzen den testu erauzketarako tresna da. Adibidez, eguraldia gaien den testuetan *eguzki* eta *tenperatura* hitzak maiztasun handiz agertzea espero da eta kirolean, aldiz, *irabazi* eta *talde*. Bestalde, *da* hitza bietan antzeko maiztasun batez agertzea espero da. Teknika honek sortzen dituen “gaiak” antzekotasuna duten hitz-multzoak dira.

2.1.4.1. Iritzi-testuen laburpena

Sentimenduen analisisiko ataza ezberdinetan, pertsona askoren iritzia jakitea beharrezkoa da; bakoitzak bere ikuspegia duelako. Pertsona askoren iritzia bildu behar izateak iritzi-testuen laburpena egin beharra eragin du. Bestalde, aspektuan oinarritutako sentimenduen sailkapenak (2.1.1 atalean erakutsi bezala) iritzi-testuen laburpen bat egiteko adina informazio eskatzen du. Egoera honetan, iritzi-testuen laburpena eta aspektu mailako sentimenduen sailkapena batzearen ondorioz, aspektuan oinarritutako iritzi-testuen laburpena sortu da. Aspektuan oinarritutako iritzi-laburpenaren emaitza bi modutakoa izan daiteke: i) testu egituratua (2.2 irudian bezala) edota ii) testu desegituratua.

Digital Camera 1:

Aspect: **GENERAL**

Positive: 105 <individual review sentences>

Negative: 12 <individual review sentences>

Aspect: **Picture quality**

Positive: 95 <individual review sentences>

Negative: 10 <individual review sentences>

Aspect: **Battery life**

Positive: 50 <individual review sentences>

Negative: 9 <individual review sentences>

...

2.2 irudia: Aspektuan oinarritutako iritzi-testuen laburpenaren emaitza egituratua (Cambria *et al.*, 2017).

2.2 irudian, iritzi-testu baten laburpenaren emaitza egituratua ikus daiteke. Bertan argazki-kamera bati buruzko iritzia ematen da eta beraren zenbait aspektu azaltzen dira: orokorra, argazkien kalitatea eta bateriaren bizitza. Aspektu horietako bakoitzaren eskuinetara, zenbat iritzi-testuk positiboki edo negatiboki ebaluatu duten zehazten da. Adibidez, argazkiaren kalitatea 95 iritzi-testuk positiboki baloratu dute eta 10 iritzi-testuk negatiboki.

Aspektuan oinarritutako iritzi-laburpena dagoen moduan, badago dokumentu askotako testu-laburpena ere. Carenini *et al.* (2006) eta Hu *et al.* (2017) eremu horretako lanak dira. Hala ere, iritzi-testuen laburpena ezberdina da dokumentu baten edo askoren laburpenetik. Izan ere, iritzi-testuen laburpe-

na aspektu eta entitateetan eta aspektu eta entitate horietako bakoitzaren orientazio semantikoan oinarritzen da. Gainera, alderdi kuantitatiboa ere badago, hau da, horietako bakoitza positiboa edo negatiboa den zehaztu behar da. Testuen laburpen tradizionalan, aldiz, esaldirik garrantzitsuenak erauzten dira eta dokumentu askotako testu-laburpenean, berriz, errepikatzen ez den informazioa bilatu eta errepikatzen den informazioa ezabatzen da.

Aspektuan oinarritutako iritzi-laburpenaren lanik aipagarrienak Hu eta Liurenak (2004) eta Liu *et al.*renak (2005) dira. Dokumentu askotako iritzi-laburpenean, aldiz, oinarritzko lana Das eta Martinsena (2007) da.

2.1.4.2. Konparaziozko iritzien analisisa

Entitate bati edo bere aspektuei buruzko iritzi positiboa edo negatiboa zuzenean adierazteko aukeraz gain, antzeko bi entitateen arteko konparazioa eginez ere iritzia adieraz daiteke. Horrelako iritziei konparaziozko iritziak deritze.

Konparaziozko esaldiek beti ez dute zuzenean iritzia adierazten. Hau da, objektu bati buruzko iritzia zuzenean eman beharrian, beste objektu batetik konparazioa da iritzia. Gainera, konparaziozko iritzien barruan, superlatibozko iritziak ere topa ditzakegu.

2.8 taulan, konparaziozko iritzietan aipatuko ditugun lanak laburbilduta daude.

Teknika	Lana
Gako-hitzak + SVM	Jindal eta Liu (2006)
Konparaziozko esaldiak identifikatzeko erregelak	Liu <i>et al.</i> (2010)
Lexikoian oinarritutako hurbilpena + aspektuan oinarritutako sentimenduen sailkapena	Ding <i>et al.</i> (2009)

2.8 taula: Konparaziozko iritziak lantzeko zenbait teknika.

Konparaziozko iritziak identifikatzeko hainbat teknika ezberdin erabili izan dira. Jindal eta Liuk (2006) erakusten dutenez, konparaziozko iritziek gako-hitzak eduki ohi dituzte, mota horretako hitzak antzematen laguntzen dutenak. Adibidez, konparaziozko adjektiboak (*gehiago*, *gutxiago*) eta superlati-

bozko adjektiboak (*gehien, gutxien*) gako-hitzak dira.

Bestalde, Liu *et al.*ek (2010) konparaziozko galderak eta konparatzen diren entitateak ikertzen dituzte. Konparaziozko esaldiak identifikatzeko erregelak erabiltzen dituzte.

Azkenik, Ding *et al.*ek (2009) konparaziozko iritzi batean entitate gustukoen identifikatzea dute helburu eta bi urrats burutzen dituzte: i) entitatea ezagutzea eta ii) entitatea esleitzea.

2.1.4.3. Iritzi-bilaketa eta -berreskuraketa

Iritzi-bilaketa web-bilaketaren atazaz eta sentimenduen analisisiko atazaz osatuta dago:

- 1- Web-bilaketako ataza: i) bilaketarekiko garrantzitsuak diren dokumentuak eta esaldiak bilatu behar dira eta, ondoren, ii) esaldi edo dokumentuak garrantziaren arabera sailkatu behar dira.
- 2- Sentimenduen analisisiko ataza: i) bilatutako dokumentu edo esaldiak bilaketa-gaiari (entitatea edo aspektua) buruzko iritzia baduen zehaztu behar da eta, hala bada, ii) iritzia positiboa edo negatiboa den zehaztu behar da.

Iritzi-bilaketan aipatu ditugun teknika ezberdinak 2.9 taulan esplikaturik daude.

Teknika	Lana
Desanbiguazioa, sinonimoen bilaketa, dokumentuen antzekotasuna + SVM	Zhang <i>et al.</i> (2007)
Sorkuntzazko hizkuntza-eredua	Eguchi eta Lavrenko (2006)
Lexikoian oinarritutako iritzi-bilaketa	Na <i>et al.</i> (2009)
Lexikoen eta sentimenduen ezaugarriak + algoritmo ezberdinak	Liu <i>et al.</i> (2009)

2.9 taula: Iritzi-bilaketa egiteko aipatutako teknikak.

Iritzi-bilaketa egiteko zenbait teknika proposatu izan dira. Zhang *et al.*ek (2007) lehenik eta behin, bilaketarekiko garrantzitsuak diren dokumentuak

bilatzen dituzte. Horretarako, kontzeptuen desanbiguazioa, kontzeptu horien sinonimoen bilaketa eta dokumentuen antzekotasuna neurtzen dituzte eta, ondoren, SVM erabiltzen dute sentimenduen sailkapena egiteko.

Eguchi eta Lavrenkok (2006), berriz, sorkuntzazko hizkuntza modelatzen¹⁸ dute iritzi-bilaketa egiteko. Horretarako, bilaketako hitz multzo adierazgarri bat, intereseko sentimendu-balentziadun hitzak erabiltzen dituzte.

Na *et al.*ek (2009), ordea, ikuspegi ezberdina lantzen dute; izan ere, lexikoian oinarrituta egiten baitute iritzi-bilaketa. Domeinuari loturiko lexikoa sortzen dute horretarako, eta lexikoia oinarrian *feedback* bidezko ikasketa dago. Ikasketa egiteko, bilaketa bateko dokumentuak erabiltzen dituzte.

Azkenik, Liu *et al.*ek (2009) iritziak dauden blog batetik algoritmo ezberdinak eta lexikoen eta sentimenduen ezaugarriak ikasten dituzte. Modu horretan, sentimenduen analisia eta osagaien berreskuraketa uztartzen dituen estrategia garatzen dute.

2.1.4.4. *Spamak* diren iritzien detekzioa

Norbanakoek eta erakundeek sare sozialetako iritziak gero eta gehiago hartzen dituzte kontuan. Iritzi horiek produktu baten erosketan, hauteskuntzeetan, marketinean edota produktu baten diseinuan eragin dezakete. Testuinguru honetan, iritzi positiboak arrakasta eta irabaziekin lotzen dira eta horregatik, norbanakoek gezurrezko iritziak eta iruzkinak idazten dituzte; produktu, zerbitzu, pertsona edo ideiei ospea emateko edo ospea kentzeko. Horrelako iritzi-motak antzemateko *spamak* diren iritzien detekzioa garatu da. Li *et al.* (2011), Lim *et al.* (2010) eta Wang *et al.* (2012) dira lanik aipagarrienak iritzi *spamak* diren iritzien detekzioan.

2.1.4.5. Iritzien kalitatearen neurketa

Sentimenduen analisiaren beste ataza bat iritzien kalitatearen neurketa da. Ataza honek iritzi-kritika baten kalitatea, lagungarritasuna, erabilgarritasuna edo baliagarritasuna neurtu nahi du. Lehenik eta behin, iritzi-testuak erauzi egiten dira eta, ondoren, iritzi-testuak sailkatu. Sailkapena egiteko

¹⁸Modelatu esatean, hizkuntzaren eredua sortu nahi dela esan nahi da.

irizpideak bere kalitatea eta bere erabilgarritasuna dira.



JOAQUIN

★★★★★ **Muy buena compra**

12 de julio de 2018

La llevo utilizando ya un tiempo y estoy encantado con ella. Muy buena calidad a ISOs altos muy buen enfoque incluso con objetivo con apertura muy grande de 1.4 o 2.8 que con mi anterior cámara EOS 700d de 5 fotos solo una salida enfocada y está la esclava todas. La pantalla táctil y abatible es comodísimo. La batería es lo único que creo que dura poco pero por lo demás es fenomenal.

A 5 personas les ha parecido esto útil

2.3 irudia: Iritziaren kalitatearen neurketa Amazon webgunean.

Gaur egun, helburu horrekin, zenbait webgunek (Amazon barne) kritikaren kalitatea edo erabilgarritasunari buruzko kalifikazioa lortzen dute irakurleei horien inguruan galdetuz. 2.3 irudian esaterako, kritika hori bost pertsoneri baliagarria suertatu zaie. Hala ere, iritziaren kalitatea neurtzeak luze jo dezake eta tresna automatikoen beharra dago.

Iritzien kalitatea neurtzeko aipatuko ditugun teknikak 2.10 taulan laburbilduta daude.

Teknika	Lana
SVM erregresioa	Kim <i>et al.</i> (2006)
Ospe ezaugarriak + eduki ezaugarriak + ezaugarri sozialak + sentimendu ezaugarriak	Lu <i>et al.</i> (2010)
Hurbilpen gainbegiratu gabea	Tsur eta Rappoport (2009)

2.10 taula: Iritzien kalitatea neurtzeko zenbait teknika.

Kim *et al.*ek (2006) iritzien kalitatearen neurketa SVM erregresioa baliatuz egin dute. Horretarako, ezaugarri hauek erabili dituzte: egitura-ezaugarriak (kritikaren luzera, esaldi-kopurua, galdera- eta harridura-ikurrak, etab.), lexiko ezaugarriak (unigramak eta bigramak), ezaugarri sintaktikoak (hitzen gramatika-kategoria eta horien agerpen-kopurua), ezaugarri semantikoak (produktuaren aspektuak eta sentimendudun hitzak) eta meta-datu ezaugarriak (kritikak dituen izar kopurua).

Lu *et al.*ek (2010), ordea, sailkatze hurbilpenean, kritika lagungarriak eta ez-lagungarriak sailkatzen dituzte. Erabiltzen dituzten ezaugarriak ospe-ezaugarriak, eduki-ezaugarriak, ezaugarri sozialak eta sentimendu-ezaugarriak dira.

Azkenik, Tsur eta Rappoportek (2009) liburu-kritiken erabilgarritasuna hurbilpen gainbegiratu gabea erabiliz neurtzen dute. Lehenik, kritiketako hitzik garrantzitsuenak identifikatzen dituzte eta horiekin bektore bat osatzen dute. Ondoren, kritika bakoitza bektore bilakatzen dute aurretik zehaztutako hitzik garrantzitsuen agerpenean oinarrituz. Azkenik, kritikak sailkatu egiten dituzte.

2.2 Sentimenduen analisirako baliabideen sorkuntza

Atal honetan, sentimenduen analisiko ataza ezberdinetan erabili ohi diren bi baliabideen inguruan jardungo dugu. Alde batetik, iritzi-testuen corpusen inguruan arituko gara (2.2.1 atala). Hauek aniztasun handia erakusten dute. Batzuk elebakarrak dira, besteak eleanitzak. Badaude anotatu gabekoak eta anotatutakoak ere. Eta batzuek testu-genero asko biltzen dituzten bitartean, besteek testu-genero bakarra lantzen dute. Bestetik, berriz, sentimenduen lexikoien inguruan arituko gara (2.2.2 atala). Baliabide hau sortzerakoan, bi hurbilpen nagusi daude: hiztegian oinarritutakoa eta corpusean oinarritutakoa.

2.2.1. Iritzi-testuen corpusak

Hizkuntza naturalaren prozesamenduko sentimenduen analisi-atazarako mota askotako eta helburu ezberdinetarako corpusak sortu izan dira. Leturia *et al.*ek (2014) corpus-mota hauek bereizten ditu.

- **Eremuaren arabera**

- Orokorrak. Denetariko gaiez osatutakoak dira. Hizkuntzaren erabilera-eremu guztietarako baliagarriak edo “adierazgarriak” izatea dute helburu.
- Espezializatuak. Hizkuntzaren erabilera-eremu espezializatu bateko testuak biltzen dituzte. Terminologiarako egokiak dira.

- **Erregistroaren arabera**

- Denetarikoak. Genero edo erregistro guztiak biltzen dituzten corpusak dira.
- Erregistro jakin batekoak. Testu guztiak genero edo erregistro jakin batekoak dira. Esaterako, kazetaritza eta administrazioeko testuez osatutako corpusak badaude.

- **Adierazgarritasunaren arabera**

- Orekatuak. Unibertsoetik hartutako testuez osatutako corpusak. Unibertso horren ezaugarriak eta banaketa badituzte. Corpusik objektiboenak dira.
- Ereduzkoak. Corpusaren egileak eredugarritzat jotzen dituen testuez osatuta dago. Helburu preskriptiboa dute horrelako corpusek.
- Oportunistak. Irizpide bakartzat testuak erraz biltzea duen corpus-mota da.

- **Hizkuntza-kopuruaren arabera**

- Elebakarrak. Hizkuntza bakarrez osatutako corpusak dira. Lexikografian erabiltzen dira, batez ere.
- Eleaniztunak. Hizkuntza bat baino gehiagoz osatutako corpusak dira. Itzulpen automatiko estatistikoko sistemak entrenatzeko egokiak dira.

- **Corpus eleaniztun motak lerrokatze mailaren arabera**

- Paraleloak. Hizkuntza bakoitzeko azpicorpuseko testuak elkarren itzulpenak diren corpusei esaten zaie.
- Konparagarriak. Azpicorpusek ezaugarri amankomunen bat dutenean esaten da. Testuak ez dira elkarren itzulpenak eta ez daude lerrokatuta. Amankomunean dutena generoa edo garaia izan daiteke.

Irizpide horietaz gain, badago beste bat corpusak sailkatzeko erabil daitekeena. Sailkapen hori aberastasunaren araberakoa da, hots, ea corpora etiketatuta dagoen edo ez.

Jarraian, sentimenduen analisirako sortu diren zenbait corpus deskribatu ditugu. Horretarako, aurretik aipaturiko ezaugarrietan oinarritu gara.

Sentimenduen analisirako sortu diren eta aipatu ditugun corpusak 2.11 taulan laburbilduta daude.

2.2. Sentimenduen analisirako baliabideen sorkuntza

Corpusa	Tamaina	Eremua	Erregistroa	Adieraz-garritasuna	Hizkuntza-kopurua	Lerrokatze-malla	Aberastasuna
Taboada (2008)	300 testu	Orokorra (6 gai)	Bakarra (iritziak)	Orekatua	Ingelesa		Etiketatu gabea
Refae eta Rieser (2014)	8.868 txio	-	Bakarra (txioak)	Oportunista	Arabiera		Hizkuntza maila ezberdinetan etiketatuta
Clematide <i>et al.</i> (2012)	270 esaldi	Orokorra	Denetariakoak	Orekatua	Alemana		Hizkuntza maila ezberdinetan etiketatuta + sentimendua
Shin <i>et al.</i> (2012)	8.050 esaldi		Bakarra (albisteak)		Koreera		MPQA hurbilpenaz etiketatuta
Li <i>et al.</i> (2012)	108 dokumentu 714 esaldi	Espezializatua (politika)	Bakarra (albisteak)		Alemana		Hizkuntza-phenomenoak + Sentimendua
Fernández <i>et al.</i> (2011)	300 testu 975 esaldi ~90.000 hitz	Orokorra	Bakarra (Web 2.0ko testuak)		Gaztelania Italiera Ingelesa	Konparagarria	Subjektibitate-adierazpenak (sentimendua)
Schulz <i>et al.</i> (2010)	750 iruzkin	Espezializatua	Bakarra (kritikak)	Orekatua	Ingelesa Gaztelania Alemana	Konparagarria	
Karoui <i>et al.</i> (2017)	38.262 txio		Bakarra (txioak)	Orekatua (ironiko eta ez-ironiko)	Frantsesa Ingelesa Italiera	Konparagarria	Sentimendua etiketatuta
Bond <i>et al.</i> (2016)	3.824 esaldi 74.732 hitz	Espezializatua	Bakarra (ipuinak)	Orekatua	Ingelesa Txinera Japoniera	Paraleloa	Kontzeptuak sentimenduariekin etiketatuta

2.11 taula: Sentimenduen analisirako sortu diren zenbait corpus.

Taboadaren (2008) *The SFU Review Corpus* izeneko lanetako bat da. Corpus honek *Epinions* webguneko iritzi-testuak biltzen ditu, sei gai edo kategorien ingurukoak: liburuak, autoak, ordenagailuak, sukaldeko tresneria, hotelak, filmak, musika eta telefonoa. Kategoría bakoitzak 25 iritzi positibo eta 25 iritzi negatibo ditu. Beraz, guztira 300 iritzi-testuko corpusa da. Bere ezaugarriei dagokienez, corpus orokorra dela esan daiteke, baita erregistro bakarrekoa dela ere, eta adierazgarritasunari dagokionez, orekatua dela. Bestalde, corpusa ingelesez dago eta hainbat hizkuntza mailatan etiketatuta dago: diskurtsoa RST hurbilpena erabiliz, subjektibitatearen ebaluazioa *Appraisal Theory* erabiliz eta, azkenik, ezeztapena eta espekulazioa.

Bestalde, Refaee eta Rieserek (2014) Twitter erabiltzen dute eta arabierazko 8.868 txio biltzen dituzte. Arabieraren dialekto ezberdinak biltzen dituzte ausaz egindako bilaketa batean. Beraz, corpus oportunistak da adierazgarritasunaren aldetik eta txioak bakarrik biltzen dituelako erregistro bakarrekoa da. Gainera, corpusa ezaugarri ezberdinez etiketatuta dago, hala nola, morfologia, sintaxia, semantika, estilistika¹⁹ eta seinale sozialak²⁰.

Clematide *et al.*ek (2012), berriz, corpusa alemanerako sortzen du. Corpusa guztira 270 esaldiz osatuta dago eta *DeWaC Corpus* izeneko corpusean oinarritzen da. Corpus honek webguneetako hainbat generotako testuak biltzen ditu. Corpusa geruza ezberdinez osatuta dago eta horietako bakoitza etiketatuta. Lehena esaldi mailako geruza da eta esaldi bakoitzaren objektibitatea, subjektibitatea eta polaritatea zehazten da. Bigarren geruza hitz eta sintagma mailakoa da, eta informazio subjektiboa eta faktuala dago etiketatuta. Azken geruza adierazpen maila da. Hizketa-gertaerak objektiboak edo zuzenekoak diren zehazten du. Beste modu batera esanda, egoera pribatuko ingurunea zehazten da. Beraz, corpus hau eremuari dagokionez orokorra dela esan daiteke, eta erregistroari dagokionez orekatua eta etiketatua.

¹⁹Hizkuntzalaritzan, estilistika adierazkortasuna eta literaturan estiloa eragiten duten hizkuntza erabilerak aztertzen dituen arloa da. Erretorikazko figurak eta sintaxizko egiturak lantzen dituzte.

²⁰Seinale sozialak gizarte gertakariei buruzko informazioa ematen duten seinale komunikatiboak edo informatiboak dira. Hainbat motariko informazioa eman dezakete: gizarte elkarrekintzak, gizarte-emozioak, gizarte-ebaluazioak, gizarte-jarrerak edota gizarte harremanak, besteak beste.

Li *et al.*ek (2012) albisteez osaturiko sentimenduen analisirako beste corpus bat aurkezten dute. Kasu honetan, alemanezko albiste politikoak bildu eta etiketatu dituzte. 108 dokumentuz eta 714 esaldiz osatuta dago corpusa. Guztira jarrerarekin lotura duten lau elementu anotatu dituzte. Horietako bat testu-aingura da eta jarrera adierazten duten testu-zatiak dira hitz, sintagma edota lokuzioak. Bigarren elementua helburua da, hots, jarreraren helburua edo jomuga. Anotatu beharreko hirugarren elementua jarreraren iturria da, hau da, iritzia adierazten duen pertsona. Azken elementua, berriz, laguntzailea deiturikoa da; sentimenduan eragin dezakeen hitza: ezeztapena, intentsifikatzailea edota ahultzailea.

Shin *et al.*ek (2012) koreerazko sentimenduen corpusa aurkezten dute. Corpusak albiste-artikuluak biltzen ditu eta guztira 8.050 esaldiek osatzen dute. Corpusa perspektiba anitzeko galde-erantzunen (*Multiperspective Question Answering*, MPQA) hurbilpenean oinarrituz etiketatuta dago. MPQAk esaldiaren orientazioa positiboa edo negatiboa den zehazteko aukera ematen du.

Orain arte, sentimenduen analisi-atazarako sortu diren corpus ezberdinak deskribatu ditugu. Ikusten denez, corpusek ezaugarri ezberdinak dituzte, bai bildutako testu edo esaldiei dagokienez, bai etiketatzeko moldeari dagokienez. Baina denek dute ezaugarri bat amankomunean eta hori corpusa hizkuntza batez osatuta dagoela da. Hala ere, badira zenbait corpus hizkuntza askoz osatuta daudenak, jarraian datozenak bezalakoak.

Fernández *et al.*ek (2011) aurrekoekiko zertxobait ezberdina den corpusa aurkezten dute. *EmotiBlog* izeneko corpusak Web 2.0 web-zerbitzuaren bidez sortutako testu-genero berriak biltzen ditu. Bertako adierazpen subjektiboak etiketatuta ditu corpusak. Corpusak 975 esaldi ditu. Horietatik 519 esaldi objektiboak dira eta gainerakoak, 456 esaldi subjektiboak. Hauen artean, 260 esaldi positiboak dira eta 188 esaldi, aldiz, negatiboak. Hiru gai lantzen dituzte testuek: Kioto protokoloa, Zimbabweko Mugaberen gobernu eta AEBetako 2008ko hauteskunde presidentzialak. Gai bakoitzeko 100 testu bildu dituzte. Bestalde, corpusa eleanitza da, gaztelaniazko, italierazko eta ingelesezko testuak baitaude. Hizkuntza bakoitzak 30.000 hitz inguru ditu.

Schulz *et al.*ek (2010) hizkuntza anitzeko corpus bat aurkezten dute. Corpus honek ingeleseko, gaztelaniako eta alemaneko testuak biltzen ditu eta

eskuz etiketatuta daude. Hu eta Liuen (2004) corpusean, ingelesez dagoe-na, oinarritzen dira corpora sortzeko eta produktu berberen alemanezko eta gaztelaniazko beste 500 iruzkin gehitzen dituzte. Iritzien ezaugarri hauek etiketatu dira: i) produktuen ezaugarriak, ii) horien orientazio semantikoa eta iii) horien indarra (0-3 artekoa). Kasu honetan, corpus eleanitza lerrokatze mailari dagokionez konparagarria da; izan ere, azpicorpusek amankomunean dutena produktu berberen iritziak izatea da.

Karoui *et al.*ek (2017) ere hiru hizkuntzaz osaturiko corpora sortu dute baina testu-mota ez da bera, txioz osaturiko corpus hirueleduna sortu baitute. Gainera, corpora helburu batekin sortu dute eta helburua ironia detektatzeko anotazio-irizpide bat finkatzea da. Corpuseko hiru hizkuntzak frantsesa, ingelesa eta italiarra dira. Frantseseko azpicorpusean, 2.073 txio ironiko eta 16.179 txio ez-ironiko daude. Ingeleseko azpicorpusean, 5.173 txio ironiko eta 6.116 txio ez-ironiko daude. Azkenik, italierakoan, 3.079 txio ironiko eta 5.642 txio ez-ironiko daude. Ikusten den moduan; beraz, corpora ez da paraleloa, txioak ez direlako itzulpenak; baina konparagarria bada, ezaugarri amankomuna ironia edukitzea edo ez edukitzea delako. Geruza askotariko anotazio-irizpidea baliatuta etiketatu dute corpora: ironia (esplizitua edo implizitua), ironia-kategoriak eta ironia-markatzaileak.

Azkenik, Bond *et al.*ek (2016) ere beste corpus eleanitz bat osatu dute. Corpus hau ingelesezko bi istorio laburrez osatuta dago. Aurretik aipatutako beste bi corpusekin corpus honek duen ezberdintasuna horixe da; corpus eleanitza itzulpen bidez lortu denez, corpora paraleloa dela. Anotazioa bi mailatan egin dute. Kontzeptu mailan, sentimendu positiboa edo negatiboa duten hitzak (kontzeptuak) etiketatu dituzte. *Chunk* mailan, berriz, sentimendu bera duten segidako hitz-multzoak etiketatu dituzte. Ingeleseko testuak originalak dira eta japonierara eta txinerara itzuli dituzte.

Gure iritzi-corporak lotura gehien Shin *et al.*ekin (2012), Li *et al.*ekin (2012) eta Taboadarekin (2008) du. Taboadak (2008) bezala, gure corpora orekatua da eta Shin *et al.* (2012) eta Li *et al.* (2012) lanek bezala, albisteak edo kazetaritzako testuak biltzen ditu gure corpusak.

2.2.2. Sentimenduen lexikoa

Sentimendua duten hitzei hainbat izen hartu izan dituzte: sentimendu-hitzak (*sentiment words*), iritzi-hitzak (*opinion words*) edota polaritate-hitzak (*polar words*), besteak beste. Liuk (2012) dioenez, sentimendu positiboko hitzek egoera edo kualitate desiratuak adierazten dituzte: *ederra*, *bikaina* edota *txundigarria*. Sentimendu negatiboko hitzek, aldiz, egoera edo kualitate ez-desiratuak adierazten dituzte: *txarra*, *beldurgarria* edota *pobrea*. Hitz horiek guztiek osatzen duten multzoari sentimenduen lexikoa esango diogu.

Sentimendu-hitzak bi motarikoak izan daitezke: oinarritutakoak (*base type*) eta konparaziozkoak (*comparative type*). Azken honetan, konparatiboak edo superlatiboak diren hitzak biltzen dira: *hobea*, *okerragoa*, *onena* eta *okerrena*, esaterako.

Liuren (2012) lanean sentimendu-hitzak biltzeko bi modu agertzen dira:

- 1- Hiztegiaren oinarritutako hurbilpena.
- 2- Corpusean oinarritutako hurbilpena.

Jarraian, bi hurbilpen horiek deskribatuko ditugu, tesi-lan honetan sortu dugun *Sentitegi*, euskarazko sentimenduen lexikoa, bi hurbilpen horietaz elikatu delako.

2.2.2.1. Hiztegiaren oinarritutako sentimenduen lexikoa

Sentimenduen lexikoa sortzeko hiztegia erabiltzea (Millerren (1995) *Wordnet* izeneko datu-base lexikala kasu) nahiko hurbilpen ohikoa izan da; izan ere, hiztegiak sinonimoak eta antonimoak biltzen dituzte eta horiek baliagarriak dira lexikoa sortzeko. Beraz, teknika sinpleena honako hau da: sentimendua duten hitz multzo txiki bat hartu eta, hiztegiaren sinonimo eta antonimo egituraren oinarrituta hiztegiaren egitura barrena mugitzea *bootstrapping* teknika²¹ erabiliz sentimendua duten hitz gehiago bilatzeko.

²¹Estatistikan, *bootstrapping* berlanginketa egiteko metodo bat da eta laginen banaketa egiteko erabiltzen da. Laginen datu multzo bat baliatuz, populazio baten gainean inferentzia egitean; ondoren, inferentzia hori laginen datuen berlanginketa baten bidez modelatzean eta, azkenik, berlangiketako datuekin inferentzia berri bat egitean datza.

Wordnet erabilia lexikoi bat Hu eta Liuren (2004) arabera modu honetan sor daiteke:

- 1- Lehenik eta behin, eskuz, hazi-hitz positibo edota negatiboak bildu behar dira. Hazi-hitzak datu-base lexikaetik hartutako hasierako zenbait hitz dira. Ondoren, hitz horietan oinarrituta bilaketa gehiago egiten dira datu-basean hitz gehiago aurkitzeko. Beigman Klebanov *et al.*ek (2012), esaterako, *anxiety*, *conflict* eta *improve* hitzak hazi-hitz moduan hartzen dituzte.
- 2- Ondoren, algoritmo batek hitz horien sinonimo eta antonimoak bilatzen ditu bai Wordneten, bai sareko beste hiztegieta. Aurkitutako hitzak zerrendara gehituko dira. Beigman Klebanov *et al.*en (2012) lanean, algoritmoaren emaitzak dira bilatutako *anxiousness* (<*anxiety*), *battle* (<*conflict*) eta *amend* (<*improve*) hitzak.
- 3- Azkenik, iterazioa behin eta berriz hasten da hitza bilatzen. Ezin direnean hitz gehiago bilatu, iterazioa amaitzen da.

Hiztegian oinarrituta lexikoia sortzea erraza eta azkarra da. Modu honetan sortutako lexikoiek akats batzuk eduki ditzakete, baina gero, eskuz konpon litezke. Modu honen desabantaila da sortutako lexikoia oso loturik daudela domeinuari nahiz testuinguruari. Ondorioz, ez dago domeinu askotarako balio duen sentimenduen lexikoirik.

Lan hauek guk sortu dugun sentimendu lexikoiarekin bi ikuspegitatik dute lotura. Alde batetik, denak dira hiztegia oinarrituta sortuak. Beste aldetik, berriz, hitzen sentimenduen balentzia kalkulatu da hitzek elkarren artean duten lotura semantikoa aintzat hartuta.

Sentimenduen lexikoa sortzeko hiztegia erabiltzen duten eta aipatuko ditugun lanak 2.12 taulan laburbilduta daude.

Blair-Goldensohn *et al.*ek (2008) *bootstrapping* teknika konplexuagoa erabiltzen dute. Izan ere, lehenik eta behin hazi-hitz positibo, negatibo eta neutralak biltzen ditu. Ondoren, hitz horiei pisua esleitzen diete grafo semantikoa

Teknika	Lana
<i>Wordnet + Bootstrapping</i>	Hu eta Liu (2004) Blair-Goldensohn <i>et al.</i> (2008)
Grafoan oinarritutako ikasketa gainbegiratu gabea	Rao eta Ravichandran (2009)
Markoven ausazko ibilaldia (<i>Markov random walk</i>)	Hassan <i>et al.</i> (2014)
Elkarrekiko informazio puntuala (<i>Pointwise Mutual Information, PMI</i>)	Turney eta Littman (2003) Taboada <i>et al.</i> (2011)

2.12 taula: Hiztegian oinarrituz sentimendu lexikoa sortzeko zenbait teknika.

erabiliz. Grafo²² semantiko horretan, ondoko adabegiak (*neighboring nodes*) *Wordnet*eko hitzen sinonimoak eta antonimoak dira eta ez daude hazi-hitz neutralen multzoan. Hazi-hitz neutralak sentimendu orientazioa hitz neutraletara ez zabaltzeko erabiltzen dira.

Rao eta Ravichandran (2009) lanak, berriz, hitzen polaritatea antzematea du helburu eta, horretarako, grafoan oinarritutako ikasketa erdi-gainbegiratu proposatzen dute. Beren ekarpena da sinonimia eta hiperonimia moduko erlazioak erabiltzea etiketen zabalkundearen emaitzak hobetzeko. Lanaren oinarrian *Wordnet* datu-base lexikala dago.

Hassan *et al.*en (2014) lanak, hitzen orientazio semantikoa kalkulatzeko asmoz, Markoven ausazko ibilaldia (*Markov random walk*) eredu erabiltzen du. Beren metodoak *Wordnet* datu-basean hedapena orientazio semantiko bera duten hitzen bidez egiten du lehenik eta, ondoren, hedapena orientazio semantiko ezberdineko hitzetara igarotzen da.

Bestalde, Turney eta Littmanen (2003) lanak elkarrekiko informazio puntuala (*Pointwise Mutual Information, PMI*) neurrian oinarritutako metodoa erabiltzen du. Hitz baten orientazio semantikoa kalkulatzeko, hitz batek hazi-hitz positiboekiko (*ona, bikaina, zuzena*) duen asoziazioaren indarra ken hazi-hitz negatiboekiko (*txarra, pobrea, okerra*) duen asoziazio semantikoa-

²²Grafoa datu-egitura mota bat eta, bertan, erlazioak edo loturak adierazten dira.

ren indarra²³ kalkulatu zenbatzen du. Elkarrekiko informazioa puntuala erabiliz neurtzen du hitz batek duen asoziazioaren indarra. Taboada *et al.*en (2011) SO-CAL tresnako lexikoa ere modu berean sortu da.

2.2.2.2. Corpusean oinarritutako sentimenduen lexikoa

Liuk (2012) azaltzen duenez, corpora erabiliz sortutako sentimenduen lexikoiak bi arrazoi jakinengatik sortzen dira.

- 1- Ezagunak diren sentimenduen hitzak edukita, domeinu bakarreko corpusean sentimendua duten hitzak eta horien orientazio semantikoa eza-gutzeko. Hau da, alde aurretik lortutako sentimendu-hitzak domeinu bakar bateko corpusean aplikatzen dira eta corpusetik sentimendu-hitzak lortzen dira.
- 2- Helburu orokor baterako edo domeinu askotarako sortutako sentimenduen lexikoa beste domeinu berri batera moldatzeko, domeinu bakarreko corpusean oinarrituz. Kasu honetan, aurretik badagoen lexikoi bat domeinu batera murrizten da corpus baliatuz.

Bi arrazoiak antzekoak diruditen arren, badaude ezberdintasun batzuk. Lehen kasuan, sentimendu-hitzak corpusetik lortu dira. Bigarrenengan, aldiz, alde aurretik daude sentimendu-hitzak, baina corpusaren iragazkia pasatu dute.

Corpusean oinarritutakoetan helburua lexikoa domeinu batera mugatzea bakarrik dela dirudien arren, hori egitea ez da uste bezain erraza. Izan ere, domeinuaren arabera, hitz batek bi aurkako esanahi izan ditzake eta hori tratatzea zaila da.

Corpusean oinarritutako lexikoiak erabili izan dira helburu orokorreko sentimenduen lexikoiak sortzeko, beti ere corpora oso handia eta anitza bada. Hala ere, helburu horretarako hiztegian oinarritutako lexikoiak egokiagoak dira.

²³ Asoziazioa estatistikan erabiltzen den kontzeptua da eta bi aldagaien arteko erlazioa edo lotura adierazten du. Sentimenduen analisisira mugatuz, bi hitzek elkarren artean duten lotura semantikoa adierazten du.

Corpusean oinarrituta sentimenduen lexikoiak sortzen dituzten lanak 2.13 taulan laburbilduta daude.

Teknika	Lana
Esaldi barnea: juntagailuak	Hatzivassiloglou eta McKeown (1997)
Esaldi barnea: juntagailuak Esaldi artean: aurkaritzako juntagailuak	Kanayama eta Nasukawa (2006)
Aspektua (iritziaren testuingurua)	Ding <i>et al.</i> (2008)
i) Elkarrekiko errefortzua ii) domeinuen arteko zati amankomuna	Du <i>et al.</i> (2010)
Antzekotasun distribuzionala	Wiebe eta Mihalcea (2006)

2.13 taula: Corpusean oinarrituta sortutako zenbait sentimenduen lexikoi.

Corpusean oinarritutako hurbilpenak guk sortu dugun sentimenduen lexikoiarekin duen lotura da biak direla hein batean corpus batetik sortuak edo corpus batean oinarrituak. Gure kasuan, lexikoi bat gaztelaniatik euskaratu ostean, lexikoa corpuseko domeinuetara egokitu dugu. Hau da, 2.2.2.2 atalaren hasieran aipatu dugun moduan, helburu orokor baterako edo domeinu askotarako sortutako sentimenduen lexikoa beste domeinu berri batera moldatu dugu, domeinu bakarreko corpusean oinarrituz.

Positiboa	Negatiboa
adequate, central, clever, famous, intelligent, remarkable, reputed, sensitive, slender, thriving	contagious, drunken, ignorant, lanky, listless, primitive, strident, troublesome, unresolved, unsuspecting

2.14 taula: Hatzivassiloglou eta McKeownen (1997) laneko hazihitzak.

Hatzivassiloglou eta McKeownena (1997) da hurbilpen honetan lehenengotariko lana. Bertan, corpora eta hazihitzak (2.14 taula) erabili dira corpusean sentimendua duten adjektibo gehiago bilatzeko. Lehenik juntagailuz lotuak dauden adjektiboak lortzen dituzte zenbait erregela linguistiko erabiliz: *eta, edo, baina, edo... edo... eta ez... ez...* izan dira. Ondoren, adjektibo bakoitzaren orientazio semantikoa lortu eta orientazioaren arabe-

Positibo bezala sailkatua	Negatibo bezala sailkatua
bold, decisive, disturbing, generous, good, honest, important, large, mature, patient, peaceful, positive, proud, sound, stimulating, straightforward, strange, talented, vigorous, witty	ambiguous, cautious, cynical, evasive, harmful, hypocritical, inefficient, insecure, irrational, irresponsible, minor, outspoken, pleasant, reckless, risky, selfish, tedious, unsupported, vulnerable, wasteful

2.15 taula: Hatzivassiloglou eta McKeownen (1997) sentimenduen le-
xikoia zati bat.

ra multzokatzen dituzte. Azkenik, multzo bakoitzaren maiztasunak alderatu eta maiztasun handienekoak positibo moduan etiketatzen dira (2.15 taula).

- (8) Koordinaziozkoa (“eta” esanahia gutxi gorabehera): *-te*, *-shi*, *-weni*, *-dakedenaku*, *-nominarazu*.
- (9) Kausazkoa (“-(e)lako” esanahia gutxi gorabehera): *-tame*, *-kara*, *-node*.
- (10) Aurkaritzakoa (“baina” esanahia gutxi gorabehera): *-ga*, *-kedo*, *-keredo*, *-monomo*, *-nodaga*.
- (11) *shikashi* (“hala ere”), *demo* (“baina”), *sorenanomi* (“nahiz eta”), *tadashi* (“baldin eta”), *dakedo* (“baina”), *gyakuni* (“bestela”), *tohaie* (“arren”), *keredomo* (“hala ere”), *ippou* (“bestalde”).

Lan horretatik abiatuz, Kanayama eta Nasukawaren (2006) lanak zenbait aurrerapen egin dituzte. Aurreko kontzeptua garatu dute. Esaldi barneko (*intra-sentential*) eta esaldi arteko (*inter-sentential*) kontzeptuak sortu dituzte. Lehena Hatzivassiloglou eta McKeown (1997) laneko kontzeptu bera da eta esaldi barnean kokatzen denez, erregela linguistikoak ere (*eta*, *edo...*) esaldi barnekoak dira ((8), (9) eta (10) adibideak). Esaldi artekoan, aldiz, jarraian dauden bi esaldiek ideia edo kontzeptu bera adierazten dute. Kasu honetan, bestelako erregela linguistikoak sortu behar izan dituzte: *baina*, *hala ere* ((11) adibidea). Esaldi artekoa izendatzeko sentimendu sendotasuna (*sentiment consistency*) kontzeptua erabili dute.

Ding *et al.*ek (2008) diotenez, goiko metodologia domenuari loturiko hitzak bilatzeko erabilgarria den arren, praktikan hori bakarrik ez da nahikoa. Izan ere, hitz batek aurkako bi orientazio semantiko izan ditzake bi testuinguru ezberdinetan. Esaterako, *bateriaren bizia luzea da* eta *fokuratzeak denbora luzea eskatzen du* esaldietan, *luzea* hitzak lehenengoan orientazio semantiko positiboa du eta bigarrenengan, berriz, negatiboa. Horregatik, domeinuari loturiko sentimenduen hitza eta haren balentzia ez direla nahikoa ohartzen dira eta horren aurrean, aspektua eta sentimendua duen hitzari garrantzi gehiago eman eta iritzi-testuingurua irudikatzeke bikoteak (aspektua, sentimendu hitza egitura dutenak) proposatzen dituzte. Adibidez, goiko adibidearen bikotea (*bateriaren bizia, luzea*) izango litzateke.

Aurreko lanak ezagunak diren hitzetatik abiatuz domeinu bateko corpusean sentimendua duten hitz gehiago bilatu eta horiei orientazio semantikoa esleitzeko hurbilpen eta teknikak dira. Jarraian, berriz, jada sortuta dagoen sentimenduen lexikoi orokor bat domeinu zehatz batera moldatzeko teknikak eta hurbilpenak aipatuko ditugu.

Du *et al.* (2010) lanean domeinu bateko lexikoa beste domeinu batera egokitzten dute. Horretarako, domeinu barruko dokumentu etiketatutak, domeinu barruko sentimendu-hitzak eta domeinuz kanpoko dokumentuak erabiltzen ditu algoritmoak. Bi aspektu lantzen dituzte: dokumentu bat positiboa bada, bertan dauden hitz askok ere orientazio semantiko positiboa dute eta dokumentua negatiboa bada, bertako hitz askok orientazio semantiko negatiboa dute (honek elkarrekiko errefortzuarekin, *mutual reinforcement*ekin, du lotura) eta nahiz eta bi domeinuen distribuzio ezberdina izan, beren arteko zati amankomuna identifikatzea posible da.

Wiebe eta Mihalceak (2006) corpusean oinarrituta hitz adierei subjektibitate etiketak esleitzea proposatzen dute. Bi azterketa egiten dituzte. Lehenengoan, *Wordnet*eko adierei bi pertsonen “subjektiboa”, “objektiboa”, “biak” etiketak esleitzeko duten adostasuna neurtzen dute. Bigarrenengan, hitzen adierei etiketak automatikoki esleitzen dizkie antzekotasun distribuzionala metodoan²⁴ oinarrituz.

²⁴Antzekotasun distribuzionala (ingelesez, *distributional similarity* teknika bat da. Teknika honen arabera, antzekoak diren objektu linguistikoek antzeko edukia (dokumentuetan

Gure sentimenduen lexikoa aurreko bi hurbilpenen uztarketa da. Alde bate-tik, hiztegian oinarritutakoa da, SO-CAL tresnaren gaztelaniazko lexikoian (Brooke *et al.*, 2009) oinarrituz egin baita. Beste aldetik, berriz, corpusean oinarritutakoa ere bada, itzultako lexikoa corpuseko sei domeinuetara ego-kitu baitugu eta anbiguoak diren hitzetan, corpuseko adiera eta, horrekin batera, sentimenduen balentzia hartu baititugu aintzat.

eta esaldietan, esaterako) eta antzeko testuingurua (hitzen kasuan) duten.

2.3 Balentzia-aldatzaileak

Behin lexikoiak nola lortzen diren aipatuta, testuinguruko balentzia-aldatzaileek horien balioetan nola eragiten duten azalduko dugu. Testuinguruzko balentzia-aldatzaileen lanketan erreferentziazko lana Polanyi eta Zaenenena (2006) da. Lan horretan, aurretik egin diren zenbait lan osatu gabe daudela adierazten dute eta beren analisisa proposatzen dute, jarraian azalduko dugun moduan.

Polanyi eta Zaenenek (2006) testuinguruko balentzia-aldatzaileak (ingelesez, *contextual valence shifters*) proposatzen dituzte. Beren arabera, testuak gertaerak eta egitateak deskribatzeaz gain, deskribatzen den gertaerarekiko idazleak edo parte-hartzaileak duten jarrera komunikatzen du. Horretarako, idazleak aukeraketa lexikala egiten du, baina testuaren antolakuntzak berak ere idazlearen jarrera adierazteko informazio kritikoa ematen du. Beste modu batera esanda, nahiz eta idazleak aukeratutako hitzen sentimendu balentzia jakin bat izan, testuaren antolakuntzak eta bertako elementuek hitzen sentimendu balentzia alda dezakete. Atal honetan aipatuko diren lanak 2.16 taulan laburbilduta daude.

2.3.1. Testuingurua kontuan hartzen ez duten lanak

Atal honetan, hitzen testuingurua kontuan hartzen ez dutelako osatugabetzat jo dituzten zenbait lan aipatu eta beren analisisa deskribatuko dugu.

Edmonds eta Hirstek (2002) sinonimo hurkoak eta aukeraketa lexikala erabiltzen dituzte. Zehatzago esanda, sinonimo hurkoen esanahiak eta beren arteko ezberdintasunak errepresentatzeko modeloa aurkezten dute, baita egoera jakin batean sinonimo hurko horien artean aukeraketa lexikal egokiena egiteko prozesua ere. Sinonimo hurkoak ia sinonimoak dira, baina ez guztiz. Esanahiaren aldetik oso antzekoak dira, baina ez dira berdinak eta ezin dira elkartrukatu. Denotazioan, konnotazioan, inplikazioan, enfasian edo erregistroan ezberdintzen dira. 2.17 taulan, zenbait sinonimo hurko ageri dira beren arteko ezberdintasuna zer den aipatuz.

Errepresentazioaren granularitatea erabiliz²⁵ sinonimo hurkoak multzokatu

²⁵Errepresentazioaren granularitatean (ingelesez, *representation granularity*), hitz ba-

Hitzen orientazio semantikoa testuingurua kontuan hartu gabe	Lana
Sinonimo hurkoak Denotazioa, konnotazioa, inplikazioa, enfasia, erregistroa	(Edmonds eta Hirst, 2002)
Juntagailuz elkartuta dauden adjektiboak	(Hatzivassiloglou eta McKeown, 1997)
Web bilaketa motorraren kontsulta Elkarrekiko informazio puntuala	(Turney eta Littman, 2002)
Maiztasun gutxiko hitzak, kolokazioa, adjektiboak, aditzak	(Wiebe <i>et al.</i> , 2004)
Berrikuntza: testuinguruko balentzia-aldatzaileak	
Esaldi mailako testuinguruko balentzia-aldatzaileak Diskurtso mailako testuinguruko balentzia-aldatzaileak	(Polanyi eta Zaenen, 2006)
Hitzen orientazio semantikoa testuingurua kontuan hartuta	
<i>General Inquirer</i> lexikoa, web corpora, Asoziazio kalifikazioa	(Kennedy eta Inkpen, 2006)
SVM Balentzia-aldatzaileen unigramak eta bigramak	(Kennedy eta Inkpen, 2006)
Ezaugarrien konbinaketa Testuinguruko balentzia-aldatzaileak	(Morsy eta Rafea, 2012)
Termino kontaketa Testuinguruko balentzia-aldatzaileak	(Ngoc Phu eta Thi Tuoi, 2014)

2.16 taula: Hitzen orientazio semantikoa esleitzeko testuingurua kontuan hartzen eta hartzen ez duten lanak.

Bariazio mota	Adibidea
Abstrakzio dimentsioa	<i>seep : drip</i>
Enfasia	<i>enemy : foe</i>
Denotaziozkoa, zeharkakoa	<i>error : mistake</i>
Denotaziozkoa, lausoa	<i>woods : forest</i>
Estilistikoa, formaltasuna	<i>pissed : drunk : inebriated</i>
Estilistikoa, indarra	<i>ruin : annihilate</i>
Jarrera adierazpena	<i>skinny : thin : slim, slender</i>
Emotiboa	<i>daddy : dad : father</i>
Kolokaziozkoa	<i>task : job</i>
Aukeraketazkoa	<i>pass away : die</i>
Azpikategorizazioa	<i>give : donate</i>

2.17 taula: Sinonimo hurkoen sailkapena beren ezberdintasunetan oinarrituta (Edmonds eta Hirst, 2002, 5. orr.).

tean esanahia sortzeko; alde batetik, testuinguruz independenteak diren esanahiek dituz-

egiten dituzte. Ondoren, ezagutza lexikalaren eredu multzokatua garatzen dute ohiko eredu ontologikoetan oinarrituz. Multzo bakoitzean azpikontzeptuak daude eta sinonimo hurkoak adierazten dituzten estiloaren, aldeko edo aurkako jarreraren, inplikazioaren eta denotazioaren arabera daude multzokatuta. Prozesua bi mailetan aritzen da aukeraketa lexikala egiteko: multzoetan eta sinonimo hurkoen multzoetan.

Hatzivassiloglou eta McKeownek (1997) adjektiboen orientazio semantikoa iragartzeko lana aurkezten dute. Horretarako juntagailuz elkartuta dauden adjektiboak baliatzen dituzte, (12), (13) eta (14) adibideetan bezalakoak.

- (12) The tax proposal was {simple and well-received} by the public.
- (13) The tax proposal was {simplistic but well-received} by the public.
- (14) The tax proposal was {*simplistic and well-received} by the public.²⁶

Erregresio linealean, (14) adibideko murriztapenak bezalakoak baliatuz (*eta* juntagailuarekin batera ezin direla orientazio semantiko negatiboko eta positiboko bi hitz agertu), adjektiboen orientazioa semantikoa iragartzea lortzen dute.

Turney eta Littmanek (2002) adjektiboen, adberbioen, izenen eta aditzen orientazio semantikoa iragartzen dute baina, horretarako, beste hurbilpen bat (metodo gainbegiratu) darabilte. Oinarritzat ehun bilioi hitzeko corpus bat, web-bilaketa motorraren kontsulta eta elkarrekiko informazio puntuala darabiltzate.

Wiebe *et al.*ek (2004) hizkuntza-subjektiboa lantzen dute, hots, iritziak, ebaluazioak edo usteak adierazteko hizkuntzaren aspektuak. Corpus batzuetatik subjektibitatearen arrastoak sortzen eta probatzen dituzte. Arrasto horiek maiztasun gutxiko hitzak, kolokazioak eta antzekotasun distribuzionala baliatuz lortutako adjektiboak eta aditzak dira. Halaber, lan horrek aipaturiko arrastoak dentsitate handi batez edo askotan hitz baten ondoan agertzeak, hitz hori subjektibo bihurtzen duela ere adierazten dute.

ten testuinguruz dependienteak diren konbinazioak eta, beste aldetik, sinonimo hurkoen ezberdintasun esplizituen multzo bat baliatzen dira.

²⁶Izartxoak (*) esaldia ez dela zuzena esan nahi du.

Laburbilduz, aipaturiko lanetan, esaldien edo testuen orientazio semantiko edo sentimendu-balentzia kalkulatzeko, hitz jakin batzuen sentimendu-balentzia erabiltzen dute. Baina, hitz horien inguruan dauden eta hitzaren sentimendu-balentzia aldatu diezaioketen hizkuntza-fenomenoak ez dira kontuan hartzen eta, horregatik, lan osatugabetzat jo izan dira.

2.3.2. Berrikuntza: testuinguruko balentzia-aldatzaileak

Polanyi eta Zaenenen (2006) arabera, aurretik aipaturiko lanek hutsune bat dute eta hutsune hori orientazio semantikoa duten hitzek ondoan duten testuingurua kontuan ez hartzea da. Izan ere, Edmonds eta Hirsten (2002) kasuan, sinonimoen arteko ezberdintasuna hitz horien zenbait ezaugarri zehaztutan oinarrituz egiten da. Hitzetik kanpo dauden ezaugarriak, hots, hitzaren testuinguruari loturikoak, ez dira aintzat hartzen. Hatzivassiloglou eta McKeownen (1997) kasuan ere, corpusetik juntagailuz loturik dauden adjektiboak ateratzen dituzte, baina ez da haien testuingurua kontuan hartzen. Turney eta Littmanen (2002) eta Wiebe *et al.*en (2004) lanekin ere berdintzatzen da, testuinguruarekin loturarik ez duten ezaugarriak bakarrik erabiltzen dira hitzaren orientazio semantikoa zehazteko.

Gure lanari dagokionez, lexikoian oinarritzen den dokumentu mailako sentimenduen sailkatzailea garatzeko bidean, jakitun gara lexikoiko sarrerek duten sentimenduen balentzia bere horretan aintzat hartuta ez dela nahikoa. Izan ere, Polanyi eta Zaenenek (2006) dioten moduan, hitz baten sentimenduen balentzia jakiteko beharrezkoa da bere testuingurua ere kontuan hartzea. Hori horrela izanik, gure lanaren eta beren lanaren arteko lotura da bietan esaldi eta diskurtso mailako balentzia-aldatzaileak identifikatzea dutela helburu. Hori dela eta, gure lanaren aurrekaria da jarraian deskribatuko dugun hurbilpena.

Osagai lexikalek jarrera positiboa edo negatiboa adierazten dute. Edozein klase irekik adieraz dezake jarrera positiboa edo negatiboa: izenak (*onarpen/katastrofe*), adjektiboak (*erraz/zail*), aditzak (*arindu/huts egin*) eta adberbioak (*azkar/mantso*) daude eta hitz anitzeko unitate lexikalak (*porrot egin*) ere izan daitezke. Horren erakusle dira (15), (16), (17) eta (18) adibideak.

- (15) 25 urteko bat bizi zen hiriko partean oinez ibili zen²⁷.
- (16) [Gizon gazte bat]₊ [bere auzoan zehar]₊ [patxadaz paseatu]₊ zuen.
- (17) [Gazte ar bat]₋ [bere lurraldean zehar]₋ [harrokeriaz ibili]₋ zen.
- (18) Filma [interesgarria]₊ eta [hunkigarria]₊ izan arren, ez zitzaidan [gustatu]₋.

(15), (16) eta (17) adibideek sentimendu-hitzen garrantzia erakusten dute. Hirurek gertaera bera azaltzen dute baina egin den aukeraketa lexikalak²⁸, hau da, aukeratutako sentimenduen balentziek ikuspegi guztiz kontrajarriak adierazten dituzte. Izan ere, (15) adibideak jarrera neutrala du, (16) adibideak positiboa eta (17) adibideak, aldiz, jarrera negatiboa. Gainera, jarrera jakin bateko hitz asko egoteak ez du esan nahi esaldiak eta testuak jarrera hori adierazten dutenik. Esaterako, (18) adibidean, jarrera positiboko hitz gehiago daude, baina esaldiak jarrera negatiboa du. Halaber, kontuan hartu behar da sentimenduen hitzek duten intentsitatea; izan ere, *ona* eta *bikain* ez baitira semantikoki berdinak. Baina Polanyi eta Zaenenek (2006) diotenez, sentimendu-hitzen ezaugarri horiek bakarrik kontuan hartzea ez da nahikoa, testuinguruko balentzia-aldatzaileek ere hiztunek bere jarrera adierazteko momentuan zerbait edo bere aurkako zerbait esatea eragin dezaketelako.

Esaldi mailan hainbat faktore (ezeztapena, aditzen tempusa, etab.) aipatzen dituzte sentimenduen balentzia alda dezaketenak.

- (19) Luar [azkarra]₊₂ da. Luar [ez da azkarra]₋₂.
- (20) Gertaera [susmagarria]₊₂ da. Gertaera [oso susmagarria]₊₃ da.
- (21) Soluzioak [eraginkorra]₊₂ →₀ izan beharko luke.
- (22) [Ana pertsona txarra da]₋₁. [Bere lagunak gaizki tratatzen ditu]₋₁. [Ana pertsona txarra izango balitz, bere lagunak gaizki tratatuko lituzke]₀.

²⁷(15) adibide hau nahiz beste hauek (16, 17, 20, 24, 25, 26, 28 eta 30) Polanyi eta Zaenenetik (2006) hartuta eta euskarara itzulita daude.

²⁸Psikolinguistikan, aukeraketa lexikala (ingelesez, *lexical selection*) hizkuntza ekoizterakoan, mezua identifikatu ondoren, erantzuteko mezua irudikatzeko hiztunak aukeratzen dituen item lexikalak dira eta, prozesu horretan, hitzari buruzko informazio semantikoa eta gramatikala aktibatzen da.

- (23) [Azterketa gainditu du.]₊ [Azterketa nekez gainditu du]_₋.
- (24) [Antolakuntza [oso [distiratsuak]₊₂]₊₃ →₋₃ [huts egin]_₋₁ zuen [arazoa konpontzen]₊₁ →₀]₊₅ →₋₄.

(19) adibidean, ezeztapenak *azkarra* adjektiboaren orientazio semantikoa aldarazten du, positibo izatetik negatibo izatera pasatzen baita. (20) adibidean, berriz, *susmagarria* adjektiboaren orientazio semantikoak bera izaten jarraitzen badu ere, balentzian²⁹ igoera bat egon da *oso* intentsifikatzailearen ondorioz. (46) adibidean, ordea, modalak esaldiaren orientazio semantikoa neutralizatu egiten du. Izan ere, deskribatzen den egoera ez da gertatu eta horrenbestez, *eraginkorra* adjektiboaren balentzia +2 izatetik 0 izatera pasatzen da. Antzeko zerbait gertatzen da (22) adibidean. Bertan, azken esaldian, tempusa alegiazkoa denez, aurreko esaldietan gertatzen den egoera zalantzan jartzen da eta ondorioz, esaldiaren orientazio semantikoa neutralizatzen da. (23) adibidean egoera ezberdina da, izan ere, *nekez* adberbioak, osagai auresuposizionala³⁰ denak, orientazio semantiko positiboko esaldia negatibo bilakatzen du. Azkenik, (24) adibidean, ironia dago eta horrek esaldiko hitzen sentimenduen balentzian aldaketak sortarazten ditu. Aldaketen ondorioz, esaldia +5 sentimendu balentzia izatetik, -4 sentimendu balentzia izatera igarotzen da.

Diskurtso mailan ere hainbat testuinguruko balentzia-aldatzaile daudela adierazten dute. Besteak beste, jarraian ageri diren adibideak aipatzen dituzte.

- (25) [Boris matematikan [aparta]₊₂ →₀ izan arren, bera oso irakasle [txarra]_₋₂ da]₀ →₋₂.
- (26) John asko ibiltzen da oinez. Aurreko hilean, astearteetan 25 milia egiten zituen oinez.
- (27) Etxea. Kokalekua. Prezioa. Hedadura.

²⁹Sentimendu-balentzia hitz batek duen informazio subjektiboa adierazteko zenbakizko eredia da. Zenbakiak eskala batean kokatuta egon ohi dira eta negatiboak edo positiboak izan daitezke hitzak duen orientazio semantikoaren arabera.

³⁰Pragmatikan, auresuposizioa (ingelesez, *presupposition*) diskurtsoan ontzat jotzen den egia duen adierazpen batekin zerikusia duen sinesmenaren edo munduaren gaineko jarrera implizitu bat da.

- (28) Hura pertsona [zerria]₋₁ da. Hark esan du hura pertsona [zerria]₋₁ →₀ dela.
- (29) Bilbo hiriaren egoera. Kokapenari dagokionez, Bilbo leku bikainean kokatzen da. Itsasoa gertu du eta inguruko mendiak ikusgarriak dira. Hiriari dagokionez, Bilbo oso kutsatuta dago eta zikinkeria edonon topa daiteke.
- (30) [Filmak [txundigarria]₊ →₀ izan beharko luke. Aktoreak [lehen mailakoak]₊ →₀ dira. Stallone gizon [zoriontsua]₊ →₀ eta [zoragarria]₊ →₀ da. Bere emaztea [gozoa]₊ →₀ [ederra]₊ →₀ da eta bere gizona [gurtzen]₊ →₀ du. Gizonak [opari]₊ →₀ [txundigarria]₊ →₀ du [bizitza]₊ →₀ [guztiz]₊ →₀ [bizitzeko]₊ →₀. Argumentuak sekulakoa₊₁ →₋₁ dirudi, hala ere, filma [porrot]₋₁ bat da]₋₂.

(25) adibidean, *izan arren* kontzesio menderagailuaren ondorioz, *aparta* adjektiboaren +2 sentimendu balentzia neutralizatu egiten da. (26) adibideak, aldiz, diskurtso-egitura bat islatzen du. Zehatz esanda, bertako bi esaldiak ELABORAZIOA izeneko erlaziozko diskurtso-egituraren bidez lotuta daude. ELABORAZIOA eta LISTA³¹ bezalako erlazioek aurretik esandakoari buruzko informazioa gehitzen dute. Ondorioz, (26) adibidean, bigarren esaldiak lehena indartzen du.

(27) adibideak, entitate askoko ebaluazio bat islatzen du. Hau da, esaldiko gai nagusia edo aspektua etxea da eta, ondoren, horri buruzko zenbait entitate daude: kokalekua, prezioa, hedadura. Bereizi egin behar dira entitateetako bakoitzak eta aspektuak berak duen orientazio semantikoa. Kasua ezberdina da (28) adibidean, bertako perpaus osagarria edo konpletiboa dago. Perpaus osagarria denez, jarreraren ebaluazioa ez du esaldia esan duenak egiten, baizik eta beste batek. Ondorioz, *zerria* hitzaren sentimendu-balentzia neutralizatu egiten da. (29) adibidean, berriz, esaldi batean hainbat azpigai daude Bilbori buruzkoak. Kokapen geografikoari dagokionez, iritziak positiboak dira, baina hiriari buruzkoak negatiboak dira. Honek lotura zuzena

³¹ELABORAZIOA erlaziozko diskurtso-egiturak aurretik aipatutako proposizio baten aspektuari buruz informazio xeheagoa ematen duen erlazioa da. LISTA erlazioan, aldiz, lotura duten proposizioak zerrendatu egiten dira.

du aspektu eta entitate mailako sentimendu-analisiarekin, izan ere, hemengo gaia eta azpigaiak hurbilpen horretako entitatea eta aspektuak dira.

(30) adibidea genero murriztapenaren adibidea da eta (27) adibidearen antzekoa da. Bertan, filmari buruzko iruzkina egiten da baina, aldi berean, filmeko pertsonaien ebaluazioa ere egiten da. Genero murriztapena dago testuan entitate bat (filma) eta bere zenbait aspektu (tartean pertsonaiak) ageri direlako. Ondorioz, filmaren iritzia positiboa edo negatiboa zehaztu nahi bada, beharrezkoa da filmeko pertsonaiei dagozkien hitzen sentimenduen balentziak kontutan hartzea.

2.3.3. Balentzia-aldatzaileetan oinarritzen diren lanak

Polanyi eta Zaenenen (2006) laneko hurbilpen teorikoa proposatu ondoren, hainbat lan daude sentimenduen sailkapena egiteko balentzia-aldatzaileak erabiltzen dituztenak.

Kennedy eta Inkpenek (2006), esaterako, filmen iruzkinen sentimenduen sailkapena egiten dute balentzia-aldatzaileak (ezeztapena, intentsifikatzaileak eta ahultzaileak) aintzat hartuz. Lehen metodoan, iruzkinak hitz-kopuru positiboak eta negatiboak zenbatuz egiten dituzte. *General Inquirer*³² bidez hitz positibo eta negatiboak eta balentzia-aldatzaileak identifikatzen dituzte eta sentimenduen lexikoi bat eta web corpusa ere badarabilte. Hitzen sentimenduen balentzia kalkulatzeko asoziazio kalifikazioa erabiltzen dute. Bigarren metodoan, berriz, euskarri bektoredun makina metodoa erabiltzen dute, balentzia-aldatzaileen unigramak eta bigramak ezaugarriak erabiliz.

Bestalde, Morsy eta Rafeak (2012) ere filmetako eta produktuetako iruzkinen sentimenduak sailkatzeko balentzia-aldatzaileak erabiltzen dituzte. Zehazki, intentsifikatzaileak, ezeztapena eta balentzia zeinua aldatzen duten sentimenduen hitzak erabiltzen dituzte ezaugarri moduan. Esperimentuetan, ezaugarri ezberdinak konbinatzen dituzte eta balentzia-aldatzaileekin eragin positiboa dutela adierazten dute.

³²*General Inquirer* (Stone *et al.*, 1966) testuko edukia ordenagailu bidez aztertzeko lexikoa da. Lexikoi honek hitzen informazio sintaktiko, semantiko eta pragmatikoa biltzen du.

Azkenik, Ngoc Phu eta Thi Tuoiek (2014) dokumentu mailako sentimenduen sailkapena egiteko ere balentzia-aldatzaileak erabiltzen dituzte. Lehenik eta behin, esaldia hitzetan segmentatzen dute. Ondoren, termino kontaketa egiten dute, hots, sentimendu-balentzia duten hitzak kontatu eta horren araberrako eragiketa egiten dute. Azkenik, sentimendu hitzei balentzia-aldatzaileen faktorea gehitzen diete eta eragiketa egiten dute berriz. Aintzat hartzen dituzten balentzia-aldatzaileak ezeztapena, konparaziozkoak eta superlatibozkoak, intentsifikatzaileak eta ahultzaileak dira.

Kasu honetan, gure lanak harreman zuzena du Ngoc Phu eta Thi Tuoiaren (2014) lanarekin. Izan ere, bietan dokumentu mailako sentimendu-sailkapena egin nahi da sentimendu lexikoi bat erabiliz. Horrez gain, bi lanetan testuinguruko balentzia-aldatzaileak ere erabili nahi dira.

2.4 Sentimenduen analisia eta euskara

Orain arte ikusi dugun moduan, sentimenduen analisia oso arlo zabala da ataza, teknika eta ikuspegi asko baitaude. Aurrekariak erakusten duten bezala, lan gehienak ingeles hizkuntzan egin dira eta, haren atzetik, beste zenbait hizkuntza handi (tartean, gaztelera, frantsesa eta txinera) daude. Egoera arrunt ezberdina da hizkuntza gutxietan eta horren arrazoi nagusietako bat baliabideen falta da. Atal honetan, euskara aintzat hartuta sentimenduen analisia arloan egin diren lanak azalduko ditugu. Lan horiek 2.18 taulan laburbilduta daude.

Lana	Ataza	Teknikak
Chen eta Skiena (2014)	Sentimendu-lexikoa	Hainbat lexikoi + jakintza-grafoa
Cruz <i>et al.</i> (2014)	Sentimendu-lexikoa	SentiWordNet 3.0 eredia
Barnes <i>et al.</i> (2018b)	Sentimendu-lexikoa	Embedding elebidunak
Saralegi <i>et al.</i> (2013)	Sentimendu-lexikoa	i) Corpusean oinarrituta ii) Hizkuntza arteko proiektzioan oinarrituta
Vicente eta Saralegi (2016)	Sentimendu-lexikoa	i) Beste hizkuntza batean oinarrituta ii) Corpusean oinarrituta ii) Jakintza base lexikalean oinarrituta
Barnes <i>et al.</i> (2018a)	Corpusaren sorkuntza	Webguneetatik iritzi-testuak lortu Hainbat geruzatan etiketatu
Vilares <i>et al.</i> (2017)	Lexikoiaren oinarritutako sentimenduen sailkapena	i) Sentimendu-lexikoa (SO-CAL eta ML-Senticon) ii) Analizatzaile sintaktikoa
San Vicente <i>et al.</i> (2015)	Aspektuan oinarritutako sentimenduen sailkapena	IXA pipe tools + Weka
Rodriguez <i>et al.</i> (2017)	Robotika	Buruaren eta besoen mugimenduak Begi mugimenduak Ahots-intonazioa Sentimendua: EliXa (IXA pipe tools + Weka)

2.18 taula: Euskara eta sentimendua uztartzen dituzten zenbait lan.

Baliabideen sorkuntzan, zehazki, sentimenduen lexikoiaren sorkuntzan, Chen eta Skienak (2014) munduko 136 hizkuntza nagusietarako, tartean euskararako, sentimenduen lexikoiak sortzen dituzte. Hainbat baliabide linguistikotan oinarrituz, jakintza-grafo handi bat sortzen dute eta hazi-hitzak zabalduz, grafoak hizkuntza bakoitzerako sentimenduen lexikoa sortzen du.

Bestalde, Cruz *et al.*ek (2014) euskara, ingelesa, gaztelania, galiziera eta katalan hizkuntzetarako sentimenduen lexikoiak sortzen dituzte. Lema mailako sentimenduen lexikoa automatikoki sortzeko metodoa aurkezten da bertan.

Beren lexikoiek hainbat geruza³³ dituzte; modu horretan, sentimendu lexikoi horiek erabiltzen dituzten aplikazioek aukera dute doitasun bat edo beste aukeratzeko eta doitasunaren arabera, lexikoia eskaintzen dituen hitz gehiago edo gutxiago aukeratzeko.

```
<?xml version="1.0" encoding="UTF-8" ?>
<layers lang="en">
  <layer level="0">
    <positive>
      <lemma pos="a" pol="1.0" std="0.0"> admirable </lemma>
      <lemma pos="a" pol="0.438" std="0.088"> amorous </lemma>
      <lemma pos="n" pol="0.594" std="0.12"> approval </lemma>
      <!-- it continues... -->
    </positive>
    <negative>
      <lemma pos="a" pol="-0.437" std="0.157"> afraid </lemma>
      <lemma pos="n" pol="-0.275" std="0.105"> aggression </lemma>
      <lemma pos="v" pol="-0.25" std="0.0"> alarm </lemma>
      <!-- it continues... -->
    </negative>
  </layer>
  <layer level="1">
    <positive>
      <lemma pos="a" pol="1.0" std="0.0"> adept </lemma>
      <lemma pos="r" pol="0.25" std="0.0"> admirably </lemma>
      <lemma pos="n" pol="0.667" std="0.315"> admiration </lemma>
      <!-- it continues... -->
    </positive>
    <negative>
      <lemma pos="v" pol="-0.875" std="0.0"> abase </lemma>
      <lemma pos="a" pol="-0.25" std="0.0"> abashed </lemma>
      <lemma pos="v" pol="-0.25" std="0.0"> abhor </lemma>
      <!-- it continues... -->
    </negative>
  </layer>
  <layer level="2">
    <!-- it continues... -->
  </layer>
</layers>
```

2.4 irudia: ML-SentiCon lexikoia ingelesezko bertsioren zati bat (Cruz *et al.*, 2014).

2.4 irudian, lexikoa geruzetan nola antolatuta dagoen ikus daiteke. Bertan, hitzak hiru geruza-mailetan (*layer level*) banatuta daude eta, horietako maila bakoitzean, hitzak daude beren sentimendu-balentziarekin. Adibidez, *amorous* (“maitekor”) hitza 0 geruza-mailan dago, adjektiboa da eta 0,088

³³Lexikoia guztira 8 geruza ditu. Lehen geruza da murriztapenik gehiena eta lema gutxi dituen eta, ondoren, geruzaz igo ahala murriztapenak gutxitu egiten dira eta lema kopurua igo egiten da. Horrela, esaterako, bigarren geruzak lehen geruzako lema eta beste berri batzuk ditu.

sentimendu balentzia du. 0 geruza-mailan dagoenez, bere doitasuna ez da oso handia.

Lema mailako lexikoa sortu aurretik, ingeleserako *synset* mailako lexikoa sortzen dute, *SentiWordNet 3.0*ren antzekoa dena. Azken hau sentimenduen lexikoirik erabiliena da gaur egun. Euskarazko sentimenduen lexikoiak 26.392 lema ditu (22.879 izen, 57 adjektibo, 3.456 aditz eta 0 adberbio) eta *ML-SENTICON* izeneko baliabidean eskuragarri dago.

Barnes *et al.*ek (2018b), berriz, besteen artean, euskarazko sentimenduen lexikoa sortzen dute; baina aurreko lanekiko hurbilpena ezberdina da. Euskara moduko baliabide gutxiko hizkuntzetan, corpus etiketatu falta handia dago eredu onak sortzeko. Horregatik, sentimendu *embedding* elebidunak (*Bilingual Sentiment Embeddings*, BLSE) sortzen dituzte beste hizkuntza bat eta lexikoa sortu nahi den hizkuntza uztartuz. Horretarako, baliabide hauek erabiltzen dituzte: i) lexikoi elebidun txiki bat; ii) jatorri hizkuntzako corpusa sentimenduz etiketatuta; iii) hizkuntza bakoitzeko hitz *embedding* elebakarrak. Esperimentuak gaztelania, katalanera eta euskaraz konbinatuz egiten dituzte esaldi mailako hizkuntzen arteko sentimenduak sailkatzean.

Saralegi *et al.*ek (2013) ere sentimenduen lexikoa sortzen dute. Euskara du aztergaitzat eta bi hurbilpen ezberdin alderatzen eta aztertzen ditu euskarazko sentimenduen lexikoiak sortzeko. Bi hurbilpenak automatikoak izateaz gain, erraz eskuratzeko baliabideetan oinarritzen dira eta, beraz, baliabide gutxiko hizkuntzetarako egokia da. Hurbilpen horietako bat corpusean oinarritutakoa da eta bestea, berriz, hizkuntzen arteko proiektzioan. Vicente eta Saralegik (2016), berriz, euskarazko sentimendu lexikoa sortzeko hiru metodo alderatzen dituzte: i) beste hizkuntza batetik itzuliz, ii) corpusetik sentimendu kalifikazioa duten hitzak erauziz eta iii) jakintza-base lexikaletik (*Lexical Knowledge Bases*) sentimenduak etiketatuz. Metodo bakoitzak zenbat eskuzko lan eta lortzen diren emaitzekiko eskuzko lan horrek merezi duten ebaluatzen dute lanean.

Bestalde, Barnes *et al.*ek (2018a) sentimenduen analisirako corpus bat sortzen dute. Bertan, aspektu mailako sentimenduen sailkapena egiteko, katalaneraz eta euskaraz dauden hotelei buruzko iritzien corpus anotatua sortzen dute. Katalanerazko 568 eta euskarazko 343 iruzkin biltzen dituzte.

Iruzkina aurreprozesatu ondoren, OpeNER proiektuko hurbilpenaz etiketatzen dituzte eta etiketatzeko *KafAnnotator* tresna erabiltzen dute. Anotatu beharreko elementuak iritzi-helburua edo jomuga, iritzi-adierazpenak, iritzi-eramailea eta polaritatea dira.

San Vicente *et al.*en (2015) *EliXa* sistema aspektuan oinarritutako sentimenduen analisia egiteko tresna da. Sistema honen abantaila da plataforma modularra dela; horrek aukera ematen du tresnari ezaugarriak kentzeko edo berriak gehitzeko. Sistema *IXA pipe tools* (Agerri *et al.*, 2014) eta Wekan oinarritzen diren hiru modulu independentez osatuta dago. Tresnari ezaugarri berriak gehitu ahal izateak egiten du posible sistema hau euskararako baliagarri izatea. *EliXa*k hizkuntza-baliabide zehatzak erabil ditzake, hala nola polaritate lexikoak eta testua normalizatzeko balio duten baliabideak. Momentuz, lau hizkuntzarako lau baliabide eskaintzen dituzte, tartean, euskararako. Euskararako baliabideak Elhuyarren sentimenduen lexikoa, maiztasun handiko hitzen zerrenda (ingelesez, *stopwords*) eta hiztegitik kanpoko hitzak (ingelesez, *Out-Of-Vocabulary*, OOV) biltzen ditu, besteak beste.

Vilares *et al.*ek (2017) sentimenduen lexikoa eta sintaxia uztartzen dituzte sentimenduen sailkapena egiteko. Iberiar penintsulako hizkuntzetarako polaritatearen sailkapena egitea dute helburu eta, horretarako, sentimenduen lexikoiak eta erregela sintaktikoez³⁴ baliatzen dira. Hurbilpen hori analizatzaile sintaktikoan oinarritzen da. Gainera, erregelak dependenteak dira eta egoera eleanizetan moldaketa bat behar dute. Euskararen kasuan ez ezik, baita katalaneran eta galizieran ere, aurretik dauden hainbat lexikoi bateratu egiten dituzte. Erabiltzen duten lexikoietakoa bat SO-CAL tresnako (Taboada *et al.*, 2011) da. Euskarazko bertsioa itzulpena eginez lortzen dute eta 4.066 sarrerako lexikoa lortzen dute. Erabilitako beste lexikoa *ML-Senticon* da. Lexikoi horren euskarazko bertsioak 1.662 sarrera ditu. Bi lexikoiak bateraketa egin ondoren, lortzen duten lexikoiak 5.134 sarrera ditu. Sortzen duten baliabidearen beste atala gramatika-kategorien etiketatzailea eta dependentsia etiketatzaileak dira. Kasu hauetan, aurretik sortuta dauden baliabideak erabiltzen dituzte. Azken urratsean, konposizio-eragiketarako egiten dituzte.

³⁴Erregela sintaktikoak esaldien egitura finkatzen dituzten erregelen multzoa da.

Rodriguez *et al.*ek (2017), berriz, aurretik aipatu dugun *EliXa* sistema robotikan aplikatzen dute. Gizakiari izaera soziala ematen diona emozioen adierazpena da. Idatzizko eta ahozko moduez gain, pertsona baten egoera mentala adierazten duen moduetako bat da. Robot baten eta gizaki baten artean, hori islatzea beharrezkoa da, interakzioa ahalik eta naturalena izateko. Lan honetan, ahozko testuari eduki emozionala jartzen diote. Buru- eta beso-mugimenduekin, begien mugimenduekin eta ahots intonazio ezberdinak erabiliz, robota gai da pozezko edo tristurako emozioak adierazteko, bai euskararen, baita ingeles eta gaztelanian ere. Robota, euskararen kasuan, emozio bat 0 eta +10 arteko eskalan sailkatzeko (bi muturrak tristura eta poztasuna dira) *EliXa* sisteman oinarritzen da.

Atal honetan aipatutako lanek badituzte zenbait berdintasun eta ezberdintasun guk garatutako tresna eta baliabideekin. Sentimenduen lexikoiaren sorkuntzan, gure hurbilpena izan da SO-CAL tresnaren (Taboada *et al.*, 2011) gaztelaniazko lexikoa euskaratzea eta ingelesekoarekin aberastea. Bestelako hurbilpenak erabili dituzte aipatu ditugun lanetan, izan ere, batzuetan *SentiWordnet* erabili dute (Cruz *et al.*, 2014), beste batzuetan, berriz, *embedding* bezalako testu-analisirako teknikak erabili dituzte (Barnes *et al.*, 2018b) eta azkenik, badira gure antzera corpusak erabili dituztenak ere (Saralegi *et al.*, 2013), nahiz eta ikuspegia ezberdina izan.

Corpusaren sorkuntzari dagokionez, Barnes *et al.*ek (2018a) gure teknika bera erabiltzen dute; hots, webguneetatik iritzi-testuak lortzen dituzte. Gero, hala ere, egindako etiketatzea ezberdina da. Azkenik, dokumentu mailako sentimenduen sailkapenean, gure lanak Vilares *et al.*enarekin (2017) antzekotasunak ditu, izan ere, bieran lexikoian oinarritutako hurbilpena eta sintaxia baliatzen dira, nahiz eta guk diskurtso maila ere lantzen dugun.

2.5 Laburpena

Kapitulu honetan, sentimenduen analisi arloan dauden ataza eta aurrekari ezberdinak aztertu ditugu. Ikusi dugunez, sentimenduen arloan hiru ataza nagusi daude: i) dokumentu mailako sentimenduen sailkapena, ii) subjektibitatearen sailkapena esaldi mailan eta iii) iritziaren jabearen erauzketa entitate eta aspektu mailan. Horietaz gain, beste ataza batzuk ere badaudela ikusi dugu eta hizkuntzaren prozesamenduaren beste arloekin lotura zuzena dutela, iritzi-laburpena edota *spam*ak diren iritzien detekzioa, adibidez.

Baliabide-sorkuntzari dagokionez, hainbat motatako iritzi-testuen corpusak daudela ikusi dugu. Testu-genero ezberdinak biltzen dituzte batzuek. Beste batzuk, berriz, corpus eleanitzak dira eta corpus gehienak etiketatuta daudela ere ikusi dugu. Sentimenduen lexikoari dagokionez, horiek sortzeko biderik ohikoena corpus edo hiztegi bidezkoa da.

Balentzia-aldatzaileen identifikazioan egin diren lanak ere landu ditugu. Balentzia-aldatzailearen kontzeptua testuaren idazleak testuak eskaintzen dituen baliabideen bidez iritzia adierazteko moldeak identifikatzeko sortu dela azaldu dugu eta, ondoren, horren inguruko erreferentziazko lana den Polanyi eta Zaenenek (2006) zer-nolako balentzia-aldatzaileak deskribatzen dituen aztertu dugu. Azkenik, balentzia-aldatzaileak kontuan hartuz sentimenduen analisisian egin diren zenbait lan aipatu ditugu.

Azken atalean, euskara eta sentimenduen analisisia lantzen dituzten lanak deskribatu ditugu. Lan gehienek euskarazko sentimenduen lexikoa sortzea dute helburutzat, askotan, horretarako SentiWordNet datu-base lexikala erabiliz.

Gure aukerei dagokienez, corpusaren kasuan, gai ezberdinetako iritzi-testuez osatuta egotea nahi dugu, baita orientazio semantikoaren aldetik orekatua eta diskurtso-egitura RST hurbilpenaz etiketatuta egotea ere. Sentimenduen lexikoari dagokionez, gure aukeraketa hiztegian eta corpusearen oinarritzen da. Guk sortu nahi dugun dokumentu maila sentimenduen sailkatzailea, berriz, lexikoian oinarritutakoa izatea hautatu dugu. Azkenik, balentzia-aldatzaileei dagokienez, Polanyi eta Zaenenek (2006) deskribatzen dituen artean, sintaxi eta diskurtso mailakoak aztertzea erabaki dugu, fonologia eta morfologia mailekoez gain.

2. SENTIMENDUEN ANALISIKO BALIABIDEAK ETA TESTUINGURUKO
BALENTZIA-ALDATZAILEAK

METODOLOGIA

3. KAPITULUA

Metodologiaren diseinua

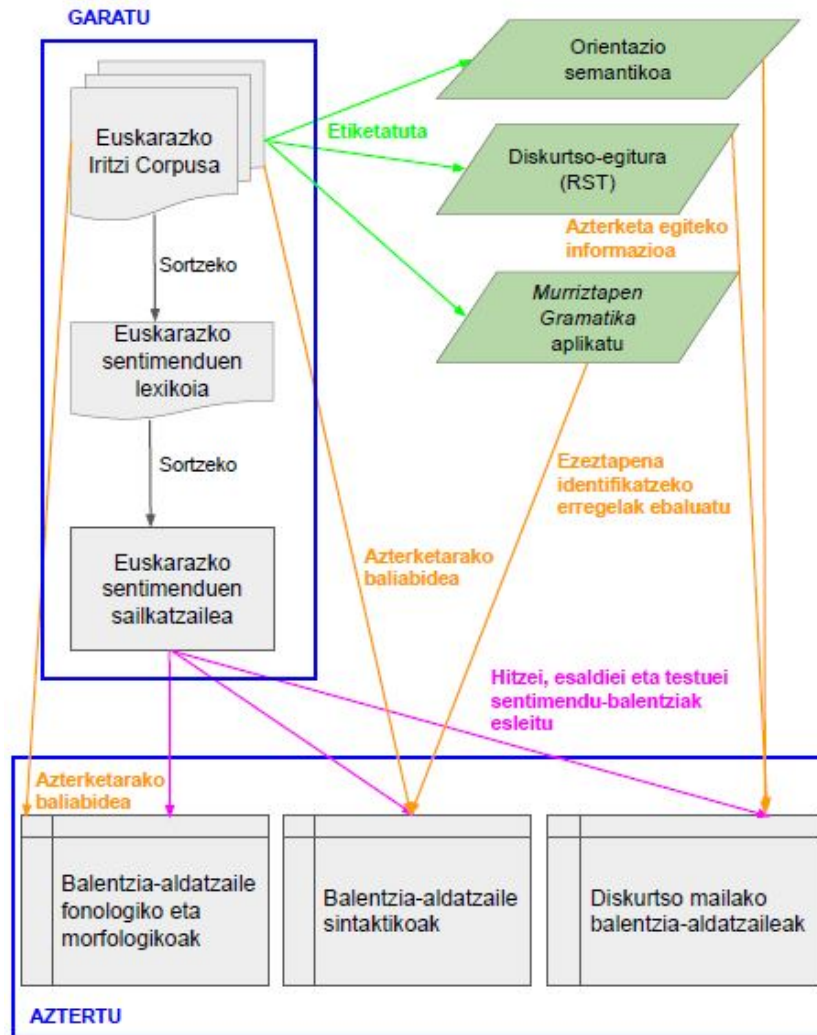
1. kapituluan zehaztutako helburuak betetzeko metodologia bat jarraitu eta zenbait baliabide eta tresna erabili ditugu. Kapitulu honetan horien inguruan arituko gara. 3.1 irudian, tesi-lan honen metodologiako urratsak laburbilduta ikus daitezke.

- 1- Ikerketa aurrera eramateko oinarritzko baliabideak sortu. Zehazki, Euskarazko Iritzi Corpusa (Alkorta *et al.*, 2016), euskarazko sentimenduen lexikoa eta euskatrazko sentimenduen sailkatzaile bat sortu ditugu.

Euskarazko Iritzi Corpusa (Alkorta *et al.*, 2016) 240 iritzi-testuz osatuta dago eta etiketatuta ere badago. Alde batetik, iritzi-testu osoaren orientazio semantikoa positiboa edo negatiboa den etiketatu dugu. Beste aldetik, corpuseko 70 iritzi-testuren diskurtso-egitura etiketatu dugu *Egitura Erretorikoaren Teoria* (RST) (Mann eta Thompson, 1988) erabiliz. Azkenik, corpuseko 48 iritzi-testuei *Murritzapen Gramatika* (MG) (Karlsson *et al.*, 1995) aplikatu diegu.

Euskarazko sentimenduen lexikoa ere sortu dugu eta, horretarako, besteak beste, Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) oinarritu gara. Azkenik, euskarazko sentimenduen sailkatzailea ere garatu dugu eta bere oinarrian sortu dugun sentimenduen lexikoa dago.

- 2- Balentzia-aldatzaileak identifikatu. Bigarren urratsa, hizkuntza maila ezberdinetan, sentimendu-balentzia duten hitzetan eragin dezaketen fenomeno linguistikoak aztertu eta beren eragina neurtzea izan da.



3.1 irudia: Tesi-lanaren metodologia.

Balentzia-aldatzaile fonologikoak aztertu ditugu eta azterketa egiteko Euskarazko Iritzi Corpusa (Alkorta *et al.*, 2016) eta euskarazko sentimenduen sailkatzailea erabili ditugu. Balentzia-aldatzaile sintaktikoe-tan ezeztapen-markak aztertu ditugu. Ezeztapen-marken adibideak Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) aurkitu ditugu eta, ondoren, ezeztapen-marka horiek identifikatzeko erregelak *Murriztapen Gramatikan* (Karlsson *et al.*, 1995) sortu eta Euskarazko Iritzi Corpu-

sari(Alkorta *et al.*, 2016) *Murriztapen Gramatika* (Karlsson *et al.*, 1995) aplikatuta, erregela horiek ebaluatu ditugu.

Azkenik, diskurtso mailan, erlaziozko diskurtso-egitura eta unitate zentrala aztertu ditugu eta balentzia-aldaketan zer-nolako eragina duten zehaztu dugu. Hori egiteko Euskarazko Iritzi Corpusa (Alkorta *et al.*, 2016) (diskurtso-egituraz eta orientazio semantikoz etiketatuta) eta euskarazko sentimenduen sailkatzailea erabili ditugu.

Jarraian, metodologiaren deskribapen zabalagoa egin dugu.

3.1 Sentimenduen analisirako baliabide eta tresnen sorkuntza

3.1.1. Euskarazko Iritzi Corpora

Euskarazko Iritzi Corpora sortzeko, lehenik eta behin, zer-nolako corpora sortu nahi dugun erabaki dugu. Corpuseko testuek jarraian azaltzen diren ezaugarriak izatea erabaki dugu:

- Iritzi-testuak balorazio argi bat izatea. Testu batzuek iritzi argirik ez dutela oharturik, aldeko edo aurkako iritzi argia duten testuak hobetsi ditugu. Iritzi argia adierazten duten iritzi-testuak biltzen baditugu, tesi-lan honetan garatutako sentimenduen sailkatzailea hobeto ebaluatuko dugu.
- Iritzi-testuak egitura sintaktiko aberatsa eta balorazioa adierazteko baliabideak edukitzea. Iritzi-testuek egitura sintaktiko ezberdinak eta balorazioa egiteko molde ezberdinak edukitzea nahi izan dugu. Modu horretan, hizkuntza mailetako eta molde ezberdinetako balentzia-aldatzaileak identifikatzeko aukera izango dugu.

Irizpideak zehaztu ondoren, euskarazko iritzi-testuak dauden webguneak bilatu ditugu. Webgune horiek espezializatuak dira edota aldizkari eta egunkariak dira. Labur azalduta, hiru iturri ezberdin erabili ditugu: i) egunkariak (hala nola, *Berrria*, *Argia*, *Naiz* eta *El Correo*), ii) webgune espezializatuak (*Kritiken Hemeroteca*, *Entzun.com*, *Zinea.eus* eta *Aizu*) eta, azkenik, iii) hainbat blog (horien artean, Joxe Landaren bloga, Baleikeko eguraldia atala eta EiTBko txirrindularitza bloga).

Euskarazko Iritzi Corpora osatzeko webguneak aurkitu ostean, corpusak eduki beharreko ezaugarriak finkatu ditugu. Ezaugarriak hiru dira:

- Corpusak tamaina handia izango du. Corpora 240 iritzi-testuz osatuta egongo da. Gure ustez, tamaina honetako corpora nahikoa da gure helburuak betetzeko eta sentimenduen analisiko beste corpusek ere antzeko tamaina dute.

3.1. Sentimenduen analisirako baliabide eta tresnen sorkuntza

3.2 irudia: Kritiken Hemerotekaren webgunea.

- Corpusa domeinu anitzekoa izango da. Iritzi-testuak sei domeinutakoak izango dira: eguraldia, politika, kirola, zinema, musika eta literatura. Corpusa domeinu askotakoa izatea nahi dugu, modu horretan, domeinuari loturiko berezitasunak edo balentzia-aldatzaile zehatzak identifikatzeko aukera izango baitugu.
- Corpuseko testuek balorazio orekatua izango dute. Domeinu bakoitza testuek erakusten duten balorazioari dagokionez orekatua izango da. Hau da, domeinu bakoitzean balorazio positiboko 20 testu eta balorazio negatiboko beste 20 testu egongo dira. Corpusa orekatua izatea erabaki dugu eredutzat hartu dugun *SFU Review Corpus* (Taboada, 2008) halakoa delako.

Hurrengo urratsean, bildutako iritzi-testuekin datu-base bat sortu dugu. Datu-base horretan, iritzi-testu bakoitzaren ezaugarri hauek zehaztu ditugu:

- Kodea. Iritzi-testu bakoitzari kode bat jarri diogu. Kode horrek beti *GAIA-zenbakia_balorazioa* egitura du.
 - Gaia: liburua (LIB), musika (MUS), pelikula (ZIN), politika (POL), kirola (KIR) eta eguraldia (EGU).

- Zenbakiak: 1etik 40 arterainoko zenbakiak jarri dizkiegu iritzi-testuei, gaiko 40 iritzi-testu bildu baititugu.
 - Balorazioa: positiboa (POS) eta negatiboa (NEG).
- Izenburua. Testu bakoitzaren izenburua izan da zehaztu dugun beste ezaugarrietako bat.
 - Helbidea. Iritzi-testua Interneten non dagoen ere zehaztu dugu, iturria onartzeko eta edozeinek lortu ahal izateko.
 - Balorazioa. Datu-basean, iritzi-testuetako bakoitza positiboa edo negatiboan den ere zehaztu dugu.
 - Tamaina. Iritzi-testuak duen hitz-kopurua ere zehaztu dugu. Hitz-kopurua kontatzeko, *Analhitza* tresna (Otegi *et al.*, 2017) erabili dugu.

Iritzi-testuetako bakoitzari loturiko informazioa datu-basean nola egituratuta dagoen ikus daiteke 3.1 taulan.

Domeinua	Kodea	Izenburua	Iturria	Balorazioa	Hitz-kopurua
Literatura	LIB29_pos	Zangotraba	Aizu	Positiboa	175
Kirola	KIR15_neg	Vigotik esku hutsik itzuli da Realá	Berria	Negatiboa	169

3.1 taula: Datu-basearen antolaketaren bi adibide.

Bestetik, iritzi-testuak formatu egokian ere jarri ditugu Hizkuntza Prozesamenduko tresnekin prozesatu ahal izateko. Testuak *txt* eta UTF-8 karaktere kodifikazio formatuan jarri ditugu.

Sortuko dugun corpusa ikerketarako egokia den ere neurtu nahi izan dugu eta horregatik, corpusa ingelesezko beste corpus subjektibo batekin (*SFU Review Corpus* (Taboada, 2008)) eta ingelesezko eta euskarazko bi corpus objektiboekin (corpus subjektiboen gai berak lantzen dituzten Wikipediako artikuluen multzoa) alderatu dugu. Honako aspektuak aztertu ditugu:

- Lehen pertsonaren presentzia. Ingelesean pertsona izenordainetan eta euskararen aditzetan, lehen pertsonak zenbateko presentzia duen neurtu dugu. Pertsonak beren esperientziak lehen pertsona singularrean edo pluralean azaltzen dituztelako aztertu dugu (Li *et al.*, 2011; Villarroel Ordenes *et al.*, 2017). Euskararen kasuan, 3.3 irudian agertzen diren instantzia guztietatik zenbat ageri diren lehen pertsona singularrean edo pluralean kontatu dugu.

```

Emaitza indibidualak
NI_ZURI 10
NR_GU 74
NR_HI 34
NK_NIK 185
NR_ZUEK 4
NR_HURA 4930
NR_HAIEK 1266
NI_ZUEI 3
NI_NIRI 77
NO 3
NI_GURI 149
NR_NI 73
NK_ZUK 22
NK_GUK 300
NI_HAIEI 77
NI_HARI 273
NK_HAIEK-K 635
NK_HIK-TO 1
NK_HIK-NO 32
NR_ZU 7
NK_HARK 2077
NK_ZUEK-K 12
NI_HIRI-NO 6

```

3.3 irudia: Euskarazko aditzetan dagoen lehen pertsonaren erabilera-
ren neurketa.

- Adjektiboen presentzia. Gramatika-kategoria guztietatik, adjektiboek zer pisu duten neurtu dugu bina corpus objektibo eta subjektibotan. Pang *et al.*ek (2002) diotenez, adjektiboak izan dira gramatika-kategoriarik erabilienak sentimenduen sailkapeneko atazetan eta doitasun ona lortu dute. Horregatik aztertu dugu adjektiboen presentzia.
- Ezeztapen-markak. Zenbait lanen arabera (Dadvar *et al.*, 2011; Wiegand *et al.*, 2010; Jia *et al.*, 2009) ezeztapen-markek hitzen edota sin-

tagmen sentimenduen balentzian eragin dezaketelako eta hori euskaraz aztertu nahi dugulako, euskarazko iritzi corpusean zenbat ezeztapen-marka dauden zenbatu dugu. Ezeztapen-marken kasuan ere, zenbaketa ordenagailuak eskaintzen dituen baliabideen bidez egin dugu.

Bukatzeko, Euskarazko Iritzi Corpora diskurtso- eta subjektibitate-informazioz etiketatu dugu. Lehenik eta behin, corpuseko 240 testuetatik 70 iritzi-testu *Egitura Erretorikoaren Teoria* (RST) hurbilpena baliatuz etiketatu ditugu. Beste lanetan etiketatu den adina testu etiketatu dugu, eta, ondorioz, ez dugu corpus osoa etiketatu. Bestalde, RST erabili dugu euskaraz gehien landu den hurbilpena delako. 3.2 taulan ikusten den moduan, guztira 70 testu (corpusaren % 29,16) etiketatu ditugu eta horietatik 19 testuek (etiketatutakoaren % 27,14) etiketatze bikoitza dute.

	A1	A2	Etiketatuak guztira	Etiketatzeko bikoitza
Zinema	30	9	30	9
Eguraldia	15	5	15	5
Literatura	5	25	25	5
Guztira	50	39	70	19

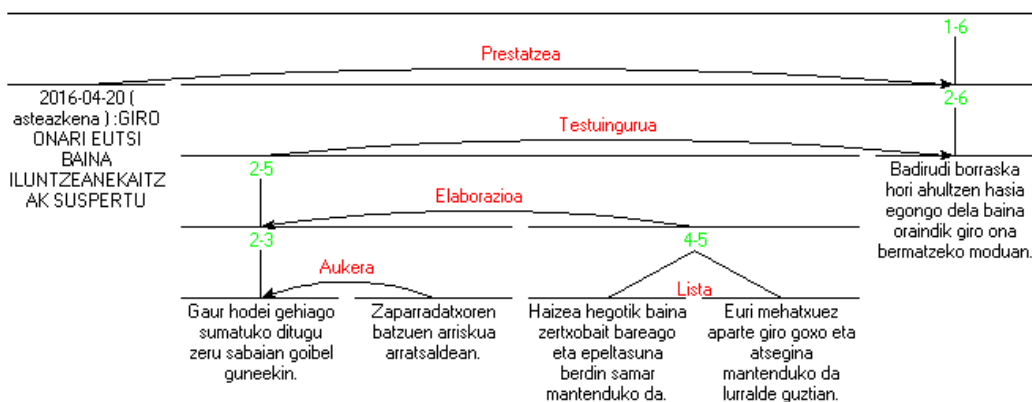
3.2 taula: Bi anotatzailek (A1 eta A2) anotatutako iritzi-testuen kopurua.

Anotazioa burutzeko, bi anotatzaileen Das eta Taboadaren (2018) irizpideei jarraitu behar izan diete. Domeinutik domeinura ezberdintasun batzuk egon dira etiketatze prozesuan. Esaterako, anotatzaileek eguraldiaren domeinuko testu bat anotatzeko 20 minutu inguru igaro behar izan dituzte eta literaturakoak anotatzeko, aldiz, ordu bete inguru. Horren arrazoiak testuen luzera eta konplexutasuna dira, eguraldikoak testu laburragoak eta sinpleagoak baitira, literaturakoaren aldean. 3.4 irudian, eguraldiaren domeinuko iritzi-testu bat (EGU04) anotatuta ikus daiteke. Bi anotatzaileen arteko adostasunaren emaitzak 3.3 taulan daude. Zehazki, erlaziozko diskurtso-egitura ondo etiketatu den neurtu dugu.

Corpusaren etiketatzea bi modutan ebaluatu dugu. Alde batetik, eskuz unitateen esanahi erretorikoa ebaluatu dugu eta, bertan, azpiosagai zentralaren¹

¹Azpiosagai zentrala *spaneko* unitaterik garrantzitsuen da (Egg eta Redeker, 2010).

3.1. Sentimenduen analisirako baliabide eta tresnen sorkuntza



3.4 irudia: EGU04 iritzi-testuaren diskurtso-egituraren anotazioa.

posizioa ez da aintzat hartu. Beste aldetik, ebaluazio kualitatiboa egin dugu. Kasu horretan, unitateen esanahi erretorikoaz gain, nuklearitatea, lotura eta osagaik hartu dira kontuan. Halaber, eskuzko ebaluazioan ez bezala, azpio-sagai zentralaren posizioa aintzat hartu da ebaluazioa egiteko momentuan.

3.3 taulako emaitzek, eskuzko ebaluazioaren emaitzek, erakusten dutenez, erlaziozko diskurtso-egitura etiketatzean bi pertsonen arteko adostasuna % 9,81ekoa da. Ezbedintasunak daude domeinutik domeinura. Eguraldiaren domeinuan, adostasuna % 43,59ra igotzen da eta zinemaren domeinuan, aldiz, adostasuna % 37,73ra jaisten da. Literaturaren domeinuan, adostasuna % 41,67koa da.

Domeinua	Adostasuna (%)
Eguraldia	43,59 (17/39)
Literatura	41,67 (70/168)
Zinema	37,73 (83/220)
Guztira	39,81 (170/427)

3.3 taula: Bi etiketatzaileen arteko adostasun-maila iritzi-testuak RST hurbilpenaz etiketatzean (eskuzko ebaluazioa).

Bi pertsonen arteko adostasuna neurtu dugu ebaluazio kualitatiboa² egiten duen tresna erabiliz ere. Emaitzak 3.4 taulan daude ikusgai.

²Iruskietak *et al.*ek (2015) aipatzen duten moduan, ebaluazio mota honi kualitatiboa

Domeinua	Osagaiak		Lotura		Nuklearitatea		Erlazioa	
	Bat egin	F1	Bat egin	F1	Bat egin	F1	Bat egin	F1
Eguraldia	20/37	0,54	9/37	0,24	22/37	0,59	15/37	0,41
Literatura	84/155	0,54	67/155	0,43	105/155	0,68	48/155	0,31
Zinema	112/221	0,56	88/221	0,40	147/221	0,67	68/221	0,31
Guztira	216/413	0,52	164/413	0,40	274/413	0,66	131/413	0,32

3.4 taula: Etiketatzailen arteko adostasunari buruz tresna automatikoak egindako ebaluazio kualitatiboa.

Eskuzko neurketan ez bezala, ebaluazio kualitatiboan³ erlaziozko diskurtso-egitura motaz gain, beste zenbait aspektu neurtu ditugu. Diskurtsoen erlazio motari dagokionez, adostasuna 0,32koa da, eskuzkoa baino 0,08 puntu baxuagoa. Ezberdintasun horren atzean, eskuzko ebaluazioan ez bezala, tresnak azpiosagai zentralaren (ingelesez, *central subconstituent*) posizioa kontuan hartzen duela dago. Ebaluatu diren beste aspektuetan, adostasuna handiagoa da. Loturaren kasuan, adostasuna 0,40koa da eta osagai eta nuklearitate aspektuetan adostasuna 0,50etik gorakoa da, 0,52 eta 0,66koak baitira, hurrenez hurren.

Diskurtso-egitura etiketatu ondoren, corpus bera subjektibitatearen ikuspegitik etiketatu da. Zehazki esanda, testuetako eta erlaziozko diskurtso-egiturako zenbait osagaien orientazio semantikoa izan da etiketatu dena.

Bi anotatzaileek erlaziozko diskurtso-egituretako hiru osagai etiketatu behar izan dituzte: i) erlaziozko diskurtso-egitura bera, ii) haren nukleoa edo nukleoak (KONTRASTEAREN erlazioaren kasuan) eta iii) haren satelitea. Esleitu beharreko etiketak ere hiru dira: i) orientazio semantiko positiboa, ii) orientazio semantiko negatiboa eta iii) orientazio semantiko neutrala.

3.5 taulan, anotatzaile bat erlaziozko diskurtso-egitura etiketatzen ikus daiteke. Anotatzaileak lehen lerroan nukleoa den testu-zati bat du eta orientazio

deritza hizkuntza ezberdinetan edo/eta pertsona ezberdinek garatutako diskurtso-egiturak alderatzeko aukera ematen duelako. Orain arteko ebaluatzeke moduak kuantitatiboak izan dira, baina honen ezberdintasuna da EDUak, testu-zatiak, nuklearitatea eta erlaziozko diskurtso-egiturak hartzen dituela kontuan.

³Ebaluazio kualitatiboak aspektu hauek aztertzen ditu: osagaia (oinarrizko diskurtso-unitateak), lotura (unitateek erlazioekin duten lotura), nukleartasuna (erlazioa N(nukleo)-S(atelite), S(atelite)-N(ukleo) edo N(ukleo)-N(ukleo) den) eta erlazioa (unitateen esanahi erretorikoa).

semantiko positiboa esleitu dio. Bigarren zutabea, satelitea den testu-zatia dago eta horri orientazio semantiko neutrala esleitu dio. Azkenik, hirugarren lerroan, erlaziozko diskurtso-egitura osoari orientazio semantiko positiboa esleitu dio.

Kodea	Erlazioa	Esaldia	Orientazio semantikoa	N-S
SENTARG01-A1.rs3	AHALBIDERATZEA	Aukera ederra da irakurlearentzat, Mirandek gaztetik hasi eta ia hil arte eutsi zion ohitura horri esker, modu kronologikoan emana gainera.	POS	N
		Froga bat egin liteke, herrien izena eta batez ere bere izena zein hizkuntzatan idazten duen begiratzea, une horretan bere herriarekiko sentimenduak eta harremanak nolakoak diren jakiteko.	NEU	S
		Aukera ederra da irakurlearentzat, Mirandek gaztetik hasi eta ia hil arte eutsi zion ohitura horri esker, modu kronologikoan emana gainera. Froga bat egin liteke, herrien izena eta batez ere bere izena zein hizkuntzatan idazten duen begiratzea, une horretan bere herriarekiko sentimenduak eta harremanak nolakoak diren jakiteko.	POS	N-S

3.5 taula: Erlaziozko diskurtso-egitura baten orientazio semantikoaren esleiketa.

Guztira, Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) dauden 240 iritzi-testuetatik 28 iritzi-testu erabili ditugu eta orientazio semantikoaren etiketatzea iritzi-testu horietako 384 erlaziozko diskurtso-egituretan egin dugu. 3.6 taulan ikusten den moduan, etiketatutako erlaziozko diskurtso-egiturek esanahi erretoriko ezberdinak dituzte⁴. Etiketatzaile batek 384 erlaziozko diskurtso-egitura etiketatu ditu eta beste batek horien % 40 etiketatu, etiketatzearen kalitatea neurtzeko. Hain zuzen ere, hori izan da etiketatzearen kalitatea neurtzeko erabili dena.

Horretarako, gidalerro batzuk sortu ditugu. Gidalerro horietan etiketatzaile bakoitzak egin beharrekoa azaltzen da. Kasu gatazkatsuenen berri ere ematen da, tartean esaldi metaforikoena, eta horiek nola anotatu behar diren ere esplikatzen da.

(31) (...) eta “egilearen intentzioa” [usna]₊ liteke. (SENTBER04)

⁴Orientazioa semantikoa etiketatuta duten iritzi-testuak *Basque RST Treebank* baliabidean (Iruskieta *et al.*, 2013) eskuragarri daude.

Erlaziozko diskurtso-egitura	Instantziak
AHALBIDERATZEA/MOTIBAZIOA	6
Baldintza azpitaldea	18
KONTRASTEIA	26
EBIDENTZIA/JUSTIFIKAZIOA	53
KONTZESIOA/ANTITESIA	70
Kausa azpitaldea	71
EBALUAZIOA/INTERPRETAZIOA	140
Guztira	384

3.6 taula: Etiketaturako erlaziozko diskurtso-egiturak.

Esate baterako, (31) adibidean, *usnatu* aditza ageri da. Ez da berak duen berezko zentzuan ageri eta, adibide horretan, *usnatu* “egileak zer egin nahi duen antzematea” esatearen parekidea da. Horregatik, mota horretako adibideen orientazio semantikoa etiketatzea erabaki dugu. Etiketatuzaileen etiketatze-adostasunaren emaitzek erakusten dutenez, Cohenen kapparen koefizientea 0,58 da. Beraz, *neurritzko adostasuna* (Landis eta Koch, 1977) da.

Bi anotatuzaileen arteko desadostasunik handienak orientazio semantiko neutralarekin lotuta daude, 3.7 taulan ikus daitezkeen moduan.

		A2			
		NEG	NEU	POS	Guztira
A1	NEG	64	27	7	98
	NEU	17	65	16	98
	POS	11	28	158	197
Guztira		92	120	181	393

3.7 taula: Erlaziozko diskurtso-egituren sentimendu balentzia anotatuari dagokion bi anotatuzaileen arteko kontingentzia-taula.

Etiketatuzaile batek orientazio semantiko neutrala esleitzen duenean, beste anotatuzaileak ez du esleitzen eta alderantziz ere gertatzen da. Hainbat etiketatze-proba egin ostean eta gidalerroak hobetu ostean, etiketatzaile batek 384 erlaziozko diskurtso-egituren gainerako % 60aren orientazio semantikoa etiketatu du.

3.1.2. *Sentitegi* izeneko euskarazko sentimenduen lexikoia- ren sorkuntza eta ebaluazioa

Euskarazko sentimendu lexikoia sortzerakoan, lehenik eta behin, horretarako zer-nolako baldintzak dauden aztertu dugu. Hiru aspektu nabarmentzen dira:

- Denbora. Sentimenduen lexikoia lanaren muina izan arren, ezin izan dugu denbora guztia horretan baliatu. Sentimenduen lexikoitik abiatuz, sentimenduen analisisa atazaren hainbat alderdi (balentzia-aldatzaileak eta dokumentu mailako sentimenduen sailkatzailea, adibidez) lantzea dugu helburu. Horrek esan nahi du denbora banatu egin behar dugula sentimenduen lexikoia sortzearen eta sentimendu-analisiko hainbat atal lantzearen artean. Ondorioz, euskarazko lexikoia sortzeko denbora mugatua dugu.
- Tresnak eta baliabideak. Lexikoia sortzeko izan dugun beste mugetako bat eskuragarri dauden euskarazko tresna eta baliabideek dituzten ezaugarriak dira. Euskarazko hainbat sentimendu lexikoi sortu izan dira hurbilpen ezberdinak baliatuta (Chen eta Skiena, 2014; Cruz *et al.*, 2014; Barnes *et al.*, 2018b; Saralegi *et al.*, 2013; Vicente eta Saralegi, 2016), baina hitzei sentimendua esleitzeko moldea guretzat ez da egokia hainbat arrazoiengatik. Batzuetan, hitzari esleitutako sentimenduen balentzia ez dago eskala batean eta, horrenbestez, ezin da hitz batean gerta daitekeen intentsitatearen aldaketa neurtu balentzia-aldatzaileen eraginez. Beste batzuetan, berriz, lexikoian, hitzen sentimenduen balentziak bi eskaletan daude (positiboan eta negatiboan) hitz horiek duten polisemia tarteko eta, azkenik, eskala bera egokia ez den kasuak ere badaude. Hori horrela izanik, sentimendu lexikoi bat euskarara itzultzea erabaki dugu.
- Kalitatea. Gure helburua ahalik eta kalitate handieneko sentimenduen lexikoia sortzea da. Aurreko puntuan aipatu bezala, tresna eta baliabideek kalitatean eragin dezakete. Bestalde, ebaluatzeko modukoa den eta gerora hobetu daitekeen lexikoi bat sortu nahi dugu. Kalitatea eta bere ezaugarriak neurtzeko, aurretik dauden lexikoen antzeko

ezaugarriak dituzten lexikoi bat eratu nahi dugu. Ondorioz, sortutako lexikoiaren kalitatea neurtzeko, dagoen lexikoi baten antzekoa sortzeko beharra ikusi dugu.

Hurrengo urratsean, aurretik aipatutako baldintzak aintzat hartuta sentimenduen lexikoa nola sortu erabaki dugu. SO-CAL tresnaren (Taboada *et al.*, 2011) gaztelaniazko lexikoa jarraian azalduko dugun metodoari jarraituz itzultzeko erabakia hartu dugu. SO-CAL tresnaren (Taboada *et al.*, 2011) ingelesezko lexikoa gaztelaniazko lexikoiko hitzari euskarazko ordaina emateko ere erabiltzea erabaki dugu. Honako abantailak ikusi ditugu, erabakia hartzerakoan:

- SO-CAL tresnako (Taboada *et al.*, 2011) lexikoen ezaugarriak. Hizkuntzako fenomeno ezberdinak kontuan hartzeko lexikoiko hitzen balentziak, sentimendua adierazten duten balioak, -5 eta +5 artean daude. Gure ustez, balentziak eta balentzien eskala egokiak dira guk sentimenduen analisisian aztertu nahi ditugun alderdietarako, hain zuzen ere, balentzia-aldatzaileek eragiten dutena eskala horretan neur baitaiteke.
- Itzulpenak egiteko baliabideak. Lehen aipatu dugun bezala, sentimenduen lexikoa euskaraz sortzeko baliabideak lortzea zaila da eta daudenak dituzten ezaugarriengatik ez dira egokiak. Baina hiztegigintzan euskaraz hainbat eta hainbat baliabide daude, eta hori dela eta, horiek erabiltzea egokia da. Besteen artean, *Elhuyar* hiztegia (Elhuyar, 2013) eta *Zehazki* hiztegia (Sarasola, 2005) sarean daude.
- Konparatzeko aukera eta ebaluazioa. Garatuko dugun sentimenduen lexikoia beste sentimenduen lexikoen ezaugarriak baditu, elkarren artean konparatzeko aukera dago eta horrek lexikoa ebaluatzeko aukera ere ematen digu.

Ondoren, sentimenduen lexikoa garatzeko baliabideak eta tresnak bildu ditugu. Guztira bost baliabide edo tresna erabili ditugu: i) SO-CAL tresnaren gaztelaniako bertsioiko lexikoa (Taboada *et al.*, 2011), ii) sarean dauden *Elhuyar* (Elhuyar, 2013) hiztegi eleanitza eta *Zehazki* (Sarasola, 2005) hiztegi

elebiduna, iii) SO-CAL tresnaren (Taboada *et al.*, 2011) ingelesezko bertsioaren lexikoa, iv) Euskarazko Iritzi Corpora (Alkorta *et al.*, 2016) eta v) KWIC izeneko teknika:

- SO-CAL tresnaren gaztelaniako bertsioiko lexikoa. Lexikoi hau izan da lanaren abiapuntua, itzuli den lexikoa baita. Guztira 4,880 sarrera ditu lexikoi honek, bost gramatika-kategorietan banatuta: izenak, adjektiboak, aditzak, adberbioak eta intentsifikatzaileak. 3.8 irudian, izenen zerrenda ikus daiteke eta bertako hitzek -5 eta $+5$ arteko sentimendu-balentzia dute.

Hitza	Sentimendu-balentzia
normalidad	2
imprudencia	-3
idiota	-3
envidia	-1
necesidad	-1
xenofobia	-4
hermosura	4
cumplimiento	3
violencia	-5
favorito	3

3.8 taula: SO-CAL tresnaren (Taboada *et al.*, 2011) gaztelaniazko bertsioaren lexikoaren zati bat.

- *Elhuyar* (Elhuyar, 2013) eta *Zehazki* (Sarasola, 2005) hiztegiak. Gaztelaniazko SO-CAL lexikoa itzultzeko sarean dauden bi hiztegiak erabili ditugu: Elhuyar⁵ eta Zehazki⁶.
- Ingelesezko SO-CAL bertsioiko lexikoa (Taboada *et al.*, 2011). Ingelesezko bertsioiko lexikoiak 6.610 hitz ditu eta bost gramatika-kategorietakoak dira: izenak, adjektiboak, aditzak, adberbioak eta intentsifikatzaileak.

⁵Elhuyar hiztegiaren esteka: https://hiztegiak.elhuyar.eus/eu_en.

⁶Zehazki hiztegiaren esteka: <http://www.ehu.eus/ehg/zehazki/>.

3.9 irudian, lexikoia zati bat ikus daiteke. Euskarazko lehen bertsioko sarrerak hiztegi honetan agertzen diren ikusi dugu eta baita zer sentimendu-balentzia duten ere. Ondoren erabaki dugu hitzari zuen sentimenduaren balentzia mantendu edo ingelesekoa esleitu.

Izena	Sentimendu-balentzia
perfection	5
beauty	4
heroism	3
compassion	2
compromise	1
accommodation	-1
complication	-2
brutality	-3
atrocitiy	-4
monstrosity	-5

3.9 taula: SO-CAL tresnaren ingelesezko bertsioren lexikoia zati bat.

- Euskarazko Iritzi Corpora (Alkorta *et al.*, 2016). Itzulpen prozesuan erabili den beste baliabideetako bat izan da. Corpusak 240 iritzi-testu biltzen ditu sei domeinutakoak: kirola, eguraldia, politika, musika, zinema eta liburuak. Corpora lexikoiko hitzen sentimenduen balentzia domeinuetara egokitzeko erabili da.
- KWIC teknika. Gaztelaniatik euskaratua izan den hitza corpusean bilatzeko eta hitzaren testuingurua ezagutzeko erabili da (ikus, 3.5 irudia). Modu horretan, gaztelaniatik euskarara itzultako hitzari domeinuari loturiko sentimenduaren balentzia esleitu diogu.

Baliabideak eta tresnak bildu ostean, sentimenduen lexikoa itzuli eta sortu dugu. Itzulpen-prozesuan hiru urrats nagusi daude: i) gaztelaniazko lexikoa euskarara itzultzea, ii) euskarazko lexikoia garbiketa eta iii) euskarazko lexikoia ebaluazioa.

3.1. Sentimenduen analisirako baliabide eta tresnen sorkuntza

Document	Left	Term ↓	Right
298) lurr...	eragiten dute. Igurtzi ezegonkor horien	ondorioz	, uhin sismikoak sortze...
298) lurr...	diren astinduak dira. Astindu horien	ondorioz	, lurrazalaren zati batzu...
293) Its...	higadura. Halaber, itsasgoren eta itsasbeheren	ondorioz	, higatutako materialak i...
290) Its...	lurrean. Bestalde, itsasgoren eta itsasbeheren	ondorioz	, higatutako materialak i...
286) IT...	atzeratuz. Bestalde, itsasgoren eta itsasbeheren	ondorioz	, higatutako materialak i...
285) Its...	sorkuntzari dagokionez, itsasgoren eta itsasbeheren	ondorioz	, higatutako materialak i...
282) Its...	Honekin batera, itsasgoren eta itsasbeheren	ondorioz	, higatutako materialak i...
282) Its...	indarrez jotzen dituzte kostaldeko haitzak;	ondorioz	, haitzak higatzen dituz...
279) Its...	lurrean. Bestalde, itsasgoren eta itsasbeheren	ondorioz	, higatutako materialak i...

ondorioz* × ? 204 context expand Scale

3.5 irudia: *Ondorioz* hitza eta bere testuinguruak.

Fenomenoa	GAZT	GAZT multzoka	EUS	ING	Azken balioa
F1	desacreditar	desacreditar -2	ospea_kendu -2 izena_kendu -2 sona_kendu -2	-	-
F2	atrofiar	atrofiar -1	atrofiatu -1	-	-
F3	amago	amago -1 cicatriz -2	seinale -1	-	-
F4	franquismo	franquismo -2	frankismo -2	-	-2
F5	correcto	acertado +3 correcto +3 decente -2	zuzen +3	right +1 correct +3	+3

3.10 taula: Itzulpen-prozesuaren azalpeneko adibideak.

Itzulpen-prozesuaren azalpena 3.10 taulako adibideetan oinarrituz egin dugu⁷.

1- Itzulpena. Urrats nagusi honetan, SO-CAL tresnaren (Taboada *et al.*, 2011) gaztelaniazko lexikoia euskarara itzuli dugu.

i) Itzulpen automatikoa gaztelaniatik euskarara. Lehenik eta behin, gaztelaniazko lexikoia automatikoki euskarara itzuli dugu *Elhuyar* (Elhuyar, 2013) eta *Zehazki* (Sarasola, 2005) hiztegiak erabiliz.

Gaztelaniazko hitz batek euskarazko ordain bat baino gehiago dituenean, ordain horietako bakoitzak hartu dugu aintzat, 3.10 tau-

⁷Euskarazko sentimenduen lexikoia sorkuntzari buruzko informazio osagarria 4.2.2 atalean dago.

lan ikus daitekeen moduan. Adibidez, F1 fenomenoan, *desacreditar* hitzaren euskarazko hiru ordain aintzat hartu ditugu: *ospea_kendu*, *izena_kendu* eta *sona_kendu*. Egoera horretan, gaztelaniazko hitzak duen balentzia; bere itzulpen-moduetako bakoitzak oinordekotzan hartu du.

- ii) Iragaketa eta multzokatzea. Urrats honetan, euskaraz lortu diren ordainak iragazi eta euskarazko ordain berdinak multzokatu egin ditugu. Hau da, gaztelaniazko lexikoiko hitzei euskarazko ordain ematean, errepikatuta dauden euskarazko ordainak batera jarri ditugu eta, beren ondoan, jatorrizko gaztelaniazko hitzak eta gaztelaniazko hitz horien sentimendu-balentziak bildu ditugu.

Horren adibide dugu, 3.10 taulan, hirugarren zutabean, F3 fenomenoan, *seinale* hitza. *Seinale* ordainaren jatorrian gaztelaniako bi hitz daude: *amago* eta *cicatriz* eta horiek, euskarazko ordain bera dutenez (ordaina errepikatuta dagoenez), multzo berean jarri ditugu, baita bere jatorrizko gaztelaniazko hitzak eta gaztelaniazko hitzen sentimendu-balentziak ere ere. F5 fenomenoko *zuzen* itzulpenaren jatorrian, berriz, gaztelaniako *acertado*, *correcto* eta *decente* daude eta euskarazko ordainak eta beren jatorrizko gaztelaniazko hitzak bildu ditugu.

- iii) *Elhuyar* (Elhuyar, 2013) edota *Zehazki* (Sarasola, 2005) hiztegi-tako sarrerak diren ordainak bakarrik aukeratu. Ordainak banan-banan aztertu ditugu eta aipaturiko bi hiztegi-tako sarrerak diren aztertu dugu. Lexikoia itzuli eta sortzerakoan, hiztegiko sarrera diren ordainak bakarrik hartuko ditugu aintzakotzat.

Esaterako, 3.10 taulan, F1 fenomenoan, *sona_kendu*, *izena_kendu* eta *ospea_kendu* ez dira hiztegi horietako sarrerak eta horrenbestez, ez ditugu aintzat hartu. *Atrofiatu*, aldiz, hiztegiko sarrera da eta kontuan hartu dugu.

- iv) Sentimenduen balentziaren aukeraketa. Hiztegi bateko sarrerak ez diren ordainak kendu ostean, geratzen diren ordainetan, euskarazko ordainari dagokion gaztelaniazko hitza eta bere sentimendu-balentzia aukeratu dizkiogu.

Euskarazko ordainari dagokion sentimenduen balentzia eta gaztelaniako esanahia aukeratzekoan, prozedura honi jarraitu diogu:

- Euskarazko ordainak itzulpen (eta balentzia) bakarra badu jatorrian gaztelaniaz, itzulpena eta hari dagokion balentzia hautatu dugu. Hori gertatu da 3.10 taulako F2 eta F4an.
- Euskarazko ordainak jatorrian hainbat gaztelaniazko hitz eta haien balentziak baditu, hitzari esleituko zaion sentimenduen balentzia (eta gaztelaniazko esanahia) Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) oinarrituta egin dugu. Hori izan da 3.10 taulako F3 eta F5en kasua.
- Kasuak egon dira non lortu den ordaina ez den corpusean agertzen eta jatorrian gaztelaniazko hainbat dituen. Egoera horretan, ordainak duen esanahirik erabiliena edo ohikoena hobetsi da.

2- Garbiketa. Bigarren urrats honetan, sentimenduen lexikoiaren lehen bertsioa garbitu egin dugu eta bigarren bertsioa sortu dugu. Zehazki esanda, lexikoia domeinu jakin batzuetara moldatu dugu eta SO-CAL tresnaren (Taboada *et al.*, 2011) ingelesezko lexikoia baliatuta aberastu eta zuzenketak egin ditugu.

v) Domeinu- eta corpus-moldaketa. Bigarren bertsioa sortzeko helburuz, lexikoiko hitzek Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) duten agerpena kontuan hartu dugu.

Adibidez, 3.10 taulan, F2 fenomenoan, *atrofiatu* Osasungintzako kontzeptu bat da eta domeinu hori ez dagoenez gure corpusean, kontzeptu hori lexikoitik ezabatu dugu.

vi) Lexikoiko sarrera bakoitza berraztertzea eta hobetzea. Lexikoiko euskarazko hitz bakoitzaren ingelesezko ordaina aurkitu dugu *Elhuyar* (Elhuyar, 2013) hiztegia erabilita eta, ondoren, hitz hori SO-CALen ingelesezko bertsioiko lexikoian agertzen den aztertu dugu. Euskarazko sarrera ingelesezko lexikoian agertzen bada, euskarazko sarrerari ingelesezko sarreraren sentimendu-balentzia esleitu diogu. Euskarazko sarrera ingelesezko bertsioiko lexikoian

ez bada agertzen, aldiz, sarrera hori kendu edo mantendu egin dugu hitzaren egokitasuna aintzat hartuz.

Esaterako, 3.10 taulan, F5 fenomenoan, euskarazko *zuzen* hitzaren baliokideak SO-CAL tresnaren (Taboada *et al.*, 2011) ingelesezko lexikoian *correct* eta *right* dira eta sentimendu-balentzia ezberdinak dituzte.

3- Lexikoiaren ebaluazioa⁸. Azken urratsean, garatutako euskarazko sentimenduen lexikoa ebaluatu dugu. Horretarako, lehendabizi, bi pertsonen anotaziotik urre-patroi bat sortu da eta ataza batean urre-patroiak eta lexikoiak ematen dituzten emaitzak alderatu ditugu.

vii) Corpuseko maiztasunik handieneko hitzen urre-patroiaren anotazioa.

Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) maiztasunik handieneko 400 hitz erauzi ditugu (gramatika-kategoria bakoitzeko 100 hitz), *Analhitza* tresna (Otegi *et al.*, 2017) erabiliz, eta bi pertsonen hitz horiei -5 eta +5 arteko sentimenduen balentzia esleitu diete.

3.11 taulan ikusten den moduan, anotatzaile batek *ahul* hitzari -3 sentimendu balentzia esleitu dio eta *argi* hitzari +5. Ondoren, bien anotazioan oinarrituz, urre-patroia eratu ditugu.

viii) Urre-patroian erabilitako 400 hitzei euskarazko sentimenduen lexikoiaren bigarren bertsioa aplikatu zaie.

ix) Urre-patroiak eta lexikoiak emandako sentimenduen balentzia esleipenak alderatu ditugu Pearson korrelazioa (Benesty *et al.*, 2009) erabiliz. Pearson korrelazioa erabiliz, adostasun neurketa bi modutan egin da:

- * Pearson 1. Neurketa honetan, 400 hitzeko zerrendan, bi pertsonen etiketatutako hitzak bakarrik hartu ditugu kontuan. Hitz bat pertsona bakar batek etiketatuta badago, hitz hori ez da kontuan hartu.

⁸Euskarazko sentimendu lexikoiaren ebaluazioari buruzko informazio gehiago 4.2.3 atalean dago.

Zenbakia	Adjektiboa	Orientazio semantikoa	Sentimendu-Balentzia
1	ageri		
2	ahul	Negatiboa	3
3	antisozial	Negatiboa	5
4	apur	Negatiboa	1
5	ar		
6	argi	Positiboa	3
7	aspergarri	Negatiboa	3
8	ausart	Positiboa	4
9	bakar	Positiboa	5
10	bakoitz		

3.11 taula: 400 hitzeko zerrendaren zati bat etiketatzaile batek bertako hitzei sentimendu-balentzia esleituta.

- * Pearson 2. Neurketa honetan, zerrendako 400 hitzak erabili ditugu. Hitzen bat pertsona batek edo bik etiketatu gabe baldin badago 0 balentzia esleitu zaio, hitz hori kontuan hartu ahal izateko.

3.1.3. Lexikoian oinarritutako dokumentu mailako euskarazko sentimenduen sailkatzailea

Lan honen hirugarren urrats nagusia lexikoian oinarritutako dokumentu mailako euskarazko sentimenduen sailkapena garatzea izan da. Helburu horrekin, SO-CAL tresnaren⁹ (Taboada *et al.*, 2011) euskarazko lehen bertsioa garatu nahi izan dugu, euskarazko sentimenduen sailkapena tresna horretan oinarritu baita. Hori dela eta, *Eustagger* tresna (Aduriz *et al.*, 2003) (3.1.3.1 atala) eta *SentiTegi* sentimenduen lexikoa (Alkorta *et al.*, 2018) (3.1.3.2 atala) erabili ditugu euskarazko bertsioa sortzeko.

⁹SO-CAL tresnaren ingelesezko bertsioaren ezaugarriak 4.3.1 atalean deskribatuta daude.

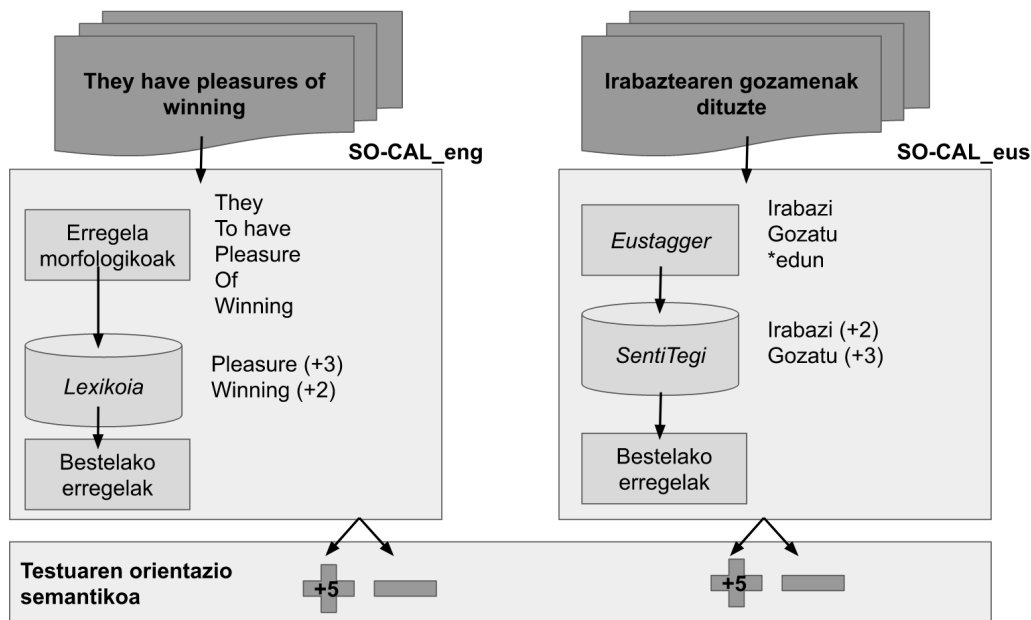
3.1.3.1. *Eustagger* tresna integratzea

Lexikoian oinarritutako sentimenduen sailkatzailea den SO-CALen (Taboada *et al.*, 2011) ingelesezko bertsioan, lematizatutako sentimendu-balentziadun hitzen lexikoa eta flexioari lotutako zenbait erregela daude. Baina euskaran, hori baliagarria izan arren, ez da modurik egokiena lexikoian oinarritutako sentimenduen sailkatzaile bat garatzeko; izan ere, tresnak testua prozesatzekotan flexioari lotutako erregela asko eduki beharko lituzke eta hori ez da bideragarria lan konplexua delako. Beste aukera bat flexioak eragindako hitz baten aldaera guztiak lexikoian sartzea izango litzateke baina, kasu horretan, lana konplexuagoa izango litzateke eta denbora gehiago eskatuko luke. Hori da euskarazko sentimenduen sailkatzailea sortzerakoan izan dugun zailtasuna.

Zailtasun horri aurre egiteko konponbidea, lexikoiak testuko hitzei sentimenduen balentzia esleitu aurretik, testuak lematizatzeko *Eustagger* tresna (Alegria *et al.*, 2002) integratzea erabaki dugu. Modu horretan, lehenik eta behin, tresnak testua lematizatuko du eta ondoren, tresnak lematizatutako hitz hori tresnaren lexikoietan badagoen aztertuko du eta baldin badago, testuko hitzari sentimenduen balentzia esleituko dio. Beraz, lematizatzaile bat tresnan integratuta, tresna gauza izango da lehenik eta behin, testua lematizatzeke; ondoren, lematizatutako hitza lexikoian dagoen aztertzeke eta, azkenik, baldin eta hitza lexikoian badago, hitzari sentimenduen balentzia esleitzeko.

Egindako aldaketa eta ingelesezko eta euskarazko SO-CAL tresnaren egituren alderaketa ageri dira, 3.6 irudian.

Ikusten den moduan, ingelesezkoan, erregela morfologikoak daude eta, modu horretan, testuetako hitzak lema bilakatzen dituzte. Euskarazkoan, aldiz, aberastasun morfologikoa tarteko, *Eustagger* lematizatzailea (Alegria *et al.*, 2002) integratu da helburu bera lortzeko asmoz. Ondoren, sentimenduen lexikoiak datoz eta lematizatutako hitzak lexikoian baldin badaude, tresnak hitz horiei sentimendu-balentzia esleitzen die. Azkenik, beste zenbait erregela daude ingeleserako, euskararako ere berak direnak (orientazio semantiko negatibodun hitzei pisu gehiago esleitzea, galde-perpausetako hitzei sentimendu balentzia ez esleitzea, etab. egiten dituztenak) eta bi hizkuntzetarako baliagarriak dira.



3.6 irudia: SO-CAL tresnaren ingelesezko eta euskarazko bertsioen egituren alderaketa.

3.1.3.2. *Sentitegi* sentimenduen lexikoia integratzea

Iritzi-testuak lematizatzeko *Eustagger* tresna (Alegria *et al.*, 2002) integratu ondoren, euskarazko sentimenduen lexikoia gauza da testuak prozesatzeko; ez, ordea, hitzei sentimendu-balentzia esleitzeko. Horretarako, tesi-lan honetan garatutako metodologiaren ondorioz sortutako *Sentitegi* euskarazko sentimenduen lexikoia (Alkorta *et al.*, 2018) integratu dugu tresna horretan.

Tresnak berak modulu bat du sentimenduen lexikoiei bideratuta; hala ingelesezko lexikoia kendu eta euskarazkoak gehitu ditugu. Tresnak funtziona dezan, beharrezkoa da lexikoia *txt* formatuan egotea eta bertan, lerro bakoitzeko sarrera bat eta bere sentimendu balentzia egotea. Hala-ber, gramatika-kategoria bakoitzak bere fitxategia eduki behar du. Izan ere, *Eustagger* lematizatzaileak (Alegria *et al.*, 2002) testuko hitza lematizatu eta bere gramatika-kategoria identifikatuko du eta ondoren, tresnak hitz hori bere gramatika-kategoriako lexikoian bilatuko du. Guk lau gramatika-kategorietako lexikoia integratu ditugu: izenena, adjektiboena, adberbioena

eta aditzena. Egin dugun aldaketa 3.12 taulan ikus daiteke.

Ingelesa		Euskara	
Thriving	+3	Bikain	+5
Record-setting	+3	Maximo	+1
Leading	+3	Orokor	+3
Industrious	+3	Min	-2
Best-selling	+3	Polit	+4
Upset	+3	Gogor	-1
Clean	+2	Ahul	-2
Capacious	+2	Behar	-1
Cogent	+2	Txar	-3
Confident	+2	Zail	-2

3.12 taula: Ingelesezko eta euskarazko sentimenduen lexikoen zati bat.

Ezkerrean, ingelesezko lexikoa ageri da, hitz-zerrenda batekin eta beren sentimenduen balentziekin. Eskuinean, berriz, euskarazko lexikoiaren zati bat dago eta horrekin ordezkatu dugu ingelesezkoa. Bi lexikoiak *txt* formatuan dauden fitxategian daude eta *SentiTegi* lexikoa integratzeko, fitxategi bat bestearekin ordezkatu dugu.

3.1.3.3. Sentimenduen sailkatzailearen ebaluazioa

Sentimenduen sailkatzailean *Eustagger* lematizatzailea (Alegria *et al.*, 2002) eta *Sentitegi* sentimendu lexikoa (Alkorta *et al.*, 2018) integratu ondoren, sailkatzailea bera ebaluatu dugu¹⁰.

Tresna ebaluatzeko Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) 48 iritzi-testu erabili ditugu, sentimenduen analisiarekin lotutako azterketa eta lanketa ez delako iritzi-testu horietan oinarrituta egin. Metodologia atal honetan aipatu bezala, corpuseko iritzi-testuen orientazio semantikoa etiketatuta dago eta sentimenduen sailkatzaileak ematen dituen emaitzak horiekin konparatu ditugu, 3.13 taulan ikusten den moduan.

¹⁰Euskarazko sentimenduen sailkatzailearen ebaluazioari buruzko informazio osagarria 4.3.3 atalean dago.

Iritzi-testua	Orientazio semantikoa	Sailkatzaileak esleitutako sentimendu-balentzia
EGU32	Positiboa	0,80 (Positiboa)
LIB30	Positiboa	1,09 (Positiboa)
ZIN39	Negatiboa	0,57 (Positiboa)

3.13 taula: Euskarazko sentimenduen sailkatzailearen ebaluazioa.

Tresnak emaitzak zenbakiz ematen ditu eta ebaluatzeko orduan, zenbakizko emaitza horretan zeinua hartu dugu kontuan, hau da, emaitza horiek zeinu positiboa edo negatiboa duten.

3.2 Balentzia-aldatzaileen identifikazioa

Ikerketa aurrera eramateko oinarritzko balibideak eta tresnak garatu ondoren, euskaran dauden testuinguruko balentzia-aldatzaileak identifikatzen hasi gara. Urratsez urrats, hizkuntza mailetakako balentzia-aldatzaileak zein diren eta nolako eragina duten aztertu dugu.

3.2.1. Fonologia eta morfologia: bustidura adierazkorra eta hizkiak

Lehenik eta behin, fonologia eta morfologia mailako balentzia-aldatzaileak identifikatzen hasi gara. Helburu horrekin, lehen urratsean, euskararen morfologia eta fonologia lantzen duten hainbat iturri bibliografikoren bilaketa egin dugu eta euskarak dituen hizkien (aurrizkiak, artizkiak eta atzizkiak) zerrenda bat osatu dugu. Zerrenda horretan, hizki bakoitza zer gramatika-kategoriarekin agertzen den, bere esanahi semantikoa zein den eta adibide bat jarri dugu. 3.14 taulako zerrenda¹¹ osatzeko erabili ditugun iturri bibliografikoen artean Mujika (1982), Euskara Institutua (2011) eta Oñederra (1990) aipa ditzakegu.

Hizkia	Izena	Adjektiboa	Aditza	Sailkapen semantikoa	Adibidea
-zale	X			Zaletasuna	Ardozale
ez-	X	X	X	Ezekotasuna	Ezberdin
-z- → -x-				Hurbiltasuna/ txikitasuna	Gazte → gaxte zahar → xahar

3.14 taula: Hizkiei buruz bildutako informazioa.

Iturri bibliografikoetatik bildutako hizki eta bustidura adierazkorren adibideak ageri dira 3.14 taulan. Lehenengo hizkia *-zale* atzizkia da, izenekin agertzen da eta zaletasuna adierazten du. Bigarrena, berriz, aurrizki bat da (*ez-*); izen, adjektibo nahiz aditzekin ager daiteke eta ezekotasuna adierazten du. Bukatzeko, azken adibidean bustidura adierazkor¹² bat dago: [z] bilakatzea

¹¹3.14 taulako zerrenda lanaren zati bat da.

¹²Bustidura adierazkorra erabilera ez-estandarrean ohikoena den arren, bere zenbait kasu aurkitu ditugu corpuseko testu formaletan.

[x]. Bustidura adierazkor guztiak bezala, ez dago murriztapen gramatikalik eta hurbiltasuna edo txikitasuna adierazteko erabiltzen da. Prozedura bera jarraitu dugu beste hizki guztiekin.

Hurrengo urratsean, Euskal Iritzi Corpuseko (Alkorta *et al.*, 2016) 192 iritzi-testu hartu ditugu eta Analhitza tresna (Otegi *et al.*, 2017) bidez aztertu ditugu. Tresna pasa ondoren, testuetako hitz guztiak maiztasunaren arabera zerrendatzea lortu ditugu. 3.15 Taulan, *Analhitza* tresnak (Otegi *et al.*, 2017) emandako hitz-zerrenda ageri da eta maiztasunean oinarritzen da. Ikusten denez, *-txo* hizkia daraman *zertxobait* hitza hamaika aldiz agertzen da corpusean.

Hitza	Instantzia-kopurua
zati	11
zertxobait	11
zerua	9

3.15 taula: Corpuseko hitzen zerrendaren zati bat maiztasunean oinarrituta.

Ondoren, euskarak dituen hizkiak eta bustidura adierazkorra egiteko moduak corpusean nola agertzen diren aztertu ditugu. Horretarako atal honetako lehen urratsean sortu dugun zerrenda erabili dugu eta, hori oinarritzat hartuz, corpusean agertzen diren hizki eta bustidura adierazkorra egiteko modu guztiekin zerrenda bat osatu dugu. Zerrendan, banan-banan, beren ezaugarriak deskribatu ditugu 3.16 taulan ikusten den moduan.

Adb.	Hitza	Hizkia	Balentzia	Balentzian eragina	Semantika	Iz.	Adj.	Adi.
(1)	istorio							
(2)	erasotxo	-txo	-2	Ahuldu	Txikitasuna	X	X	
(3)	kuttun	[t] → [tt]	-2	Indartu	Hurbiltasuna			

3.16 taula: Corpusean agertutako hizkien eta bustidura adierazkoren azterketaren lagin bat

3.16 taulan, (1) adibidean, bustidura adierazkorrik edo hizkirik ez da agertzen. (2) adibidean, *-txo* atzizkia ageri da eta honek *eraso* hitzaren sentimendu balentzia (-2) ahultzen du. (3) adibidean, azkenik, bustidura adierazkorra

dago [t] [tt] bilakatu baita eta aldaketa horrek adierazkortasuna indartu duenez, duen sentimenduaren balentzia are indartsuagoa bilakatzen da. Guztira 59 atzizki, 7 aurrizki eta 13 bustidura adierazkorraren instantzia aurkitu ditugu Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) eta horiekin guztiekin prozedura bera jarraitu dugu.

3.2.2. Sintaxi maila: ezeztapen-markak

Hitzen balentzian eragin dezakeen hizkuntzaren beste alderdi bat sintaxia da. Baina kasu honetan, balentzia aldatzearen ondorioa jasoko duena ez da hitz bat baizik eta hitz multzo batek osaturiko sintagma edo esaldi bat izan daiteke. Hau da, morfologiako eta fonologiako balentzia-aldatzaileek hitzari eragiten dioten moduan, sintaxiko balentzia-aldatzaileek hitzi ez ezik, sintagma edota esaldiari ere eragin diezaiokete.

Sintaxian, ezeztapen-markek esaldien sentimendu balentzia nola aldatzen duten aztertzeke asmoz, lehenik eta behin, euskararen ezeztapen-marken zerrenda osatu dugu eta zerrenda hori Altuna *et al.*en (2017) eta Altuna *et al.*en (1985) lanetan oinarrituta osatu dugu. Honako hauek izan dira ikerketarako erabili ditugun ezeztapen-markak: *ez*, *ezin*, *gabe*, *ezik*, *salbu*, *ezta* eta *ezean*.

Hurrengo urratsean, Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) 96 testutan, ezeztapen-markaren bat duten esaldiak bilatu ditugu eta horiekin esaldi-zerrenda bat osatu dugu, 3.17 taulan ikus daitekeen moduan. Ez ditugu testu guztiak erabili, Altuna *et al.*en (2017) lanean ageri diren ezeztapen-marka gehienak lortu ditugulako eta testu gehiago azertu arren, ezeztapen-marka berririk ez delako agertu. Guztira 359 ezeztapen-marken instantzia lortu dugu. *Analhitza* tresnak (Otegi *et al.*, 2017) emandako datuen arabera, 359 ezeztapen-marka horiek 320 esalditan banatuta daude. Horrenbestez, badaude esaldiak non ezeztapen-marka bat baino gehiago dauden. Bestalde, lortu dugun ezeztapen-marken azpicorpus honek 5.515 hitz ditu.

Ondoren, 320 esaldi horietako hitzei nahiz esaldi osoei sentimendu balentzia esleitu diegu. Hori egiteko, hitzei kategoria gramatikala edota marka lexikala esleitzen dien *Eustagger* tresnari (Aduriz *et al.*, 2003) *Sentitegi* (Alkorta *et al.*, 2018) sentimenduen hiztegiko hitzak gehitu dizkiogu; lexikoian hitzak

TESTUA	ITEMA
EGU23	Ez da espero euri asko egitea, baina giroa hezea eta freskoa izango da.
EGU23	Asko ez du egingo, baina zerbait bai.
EGU24	Zumaian ez du elurrik egingo baina elur-kota 400-500 metrora jaitsiko da, beraz kontuz errepideetan.
EGU24	Oraintxe bertan ematen du Frantzia iparraldera iritsiko dela eta hortik haize indartsua sortuko du gure lurraldean, baina ez da izango lehengokoa bezain indartsua.
EGU25	Abuztua hasi berri dugun honetan, ez dugu urte sasoi honetarako espero izan ohi dugun eguraldirik izango.
EGU25	Goizean berriz aterri mantenduko du eta ez du ia euririk egingo.

3.17 taula: Ezeztapen-markak aztertzeke sortutako azpicorpusaren zati bat.

kategoria gramatikalean oinarrituta banatuta baitaude.

- (32) Pogostkinak ezin [hobeki]₊₂ atera zituen. (MUS20)
- (33) [Irabazi]₊₂ ezinik jarraitzen du Eibarrek. (KIR17)
- (34) Ikuspuntu [politikotik]₋₁ ez ezik, [ekonomikotik]₊₃ ere Greziak esperantza ekarri du Europako hegoaldeko beste herrietara, tartean Euskal Herrira. (POL08)

(32), (33) eta (34) adibideetan, etiketatzailerak eta hari gehitutako sentimendu lexikoiak lau hitzei esleitu die sentimendu balentzia: *hobeki* adjektiboari, *irabazi* aditzari, *politiko* adjektiboari eta *ekonomiko* adjektiboari. Lehen bi hitzei +2 balentzia esleitu die eta adjektiboei -1 eta -3, hurrenez hurren. Modu honetan, ezeztapen-markak eta sentimendu balentziadun hitzak edukita, analisi linguistikoa egiteko ingurunea prest utzi dugu. Eta horrenbestez, analisi linguistikoa egitea da hurrengo urratsa.

Ezeztapen-marken analisi linguistikoa egiterakoan, aintzat hartu duguna izan

da ezeztapen-markak eragiten duen aldaketa orientazio semantikoan eta zehazki, sentimendu-balentzian. Hots, ezeztapen-markak sentimendu-balentziadun hitz eta hitz-multzoetan, haren irismenean eta ondorioz, baita esaldiko sentimendu-balentzian ere, zer ondorio uzten duen aztertu nahi izan dugu. Azterketa eskuz egin dugu eta jarraian azaltzen den prozedura jarraitu dugu.

Azterketan, corpuseko esaldiak banan-banan aztertu ditugu eskuz. Esaldietako ezeztapen-markak identifikatu ondoren, haren irismena zein den adierazi dugu. Azkenik, irismen-eremuko sentimendu-balentzian nahiz esaldi osokoan utzi duen ondorioa aintzat hartuta, esaldia bera eta haren ezeztapen-marka multzokatu egin ditugu.

Azter ditzagun, berriz, aurretik aipaturiko hiru adibideak.

- (35) Pogostkinak [ezin hobeki₊₂] atera zituen. (MUS20)
- (36) [Irabazi₊₂ ezinik] jarraitzen du Eibarrek. (KIR17)
- (37) [Ikuspuntu politikotik₋₁ ez ezik], [ekonomikotik₊₃] ere Greziak esperantza ekarri du Europako hegoaldeko beste herrietara, tartean Euskal Herrira. (POL08)

(35) adibidearen kasuan, ezeztapen-marka *ezin* da eta bere irismena *hobeki* adjektiboraino iristen da. Irismen-eremuan hitz bakarra dagoenez eta balentziaduna denez, irismen-eremuaren balentzia +2 da, baita esaldiarena ere. Kasu honetan, ezeztapen-markaren eragina aztertuta, ezeztapen-markak *hobeki* balentziadun hitza indartu egiten duela ohartu gara, intentsitatez *ezin hobeki hobeki* baino indartsuagoa baita. Ondorioz, esaldia bera eta ezeztapen-marka balentzia indartzen duten taldean sailkatu ditugu.

(36) adibidean, aldiz, nahiz eta ezeztapen-marka bera den (*ezin*), ikusi dugu haren eragina ezberdina dela. Hau da, *ezin* ezeztapen-markaren irismena *irabazi* da, +2 balentzia duena eta honi indarra ahultzen dio. Hots, *irabazi ezinik irabazi* baino ahulagoa da sentimendu-balentziari dagokionez. Ondorioz, ezeztapen-marka eta esaldia bera beste multzo batean sartu ditugu, ahultze bat egon delako. Bi adibide hauekin ikusi den moduan, posible da ezeztapen-marka berak bi funtzio izatea.

Azkenik, (37) adibidean, *ez ezik* ezeztapen-markak -1 sentimendu-balentzia

duen *ikuspuntu politikotik* hitz multzoa ezeztatzen du, baina sentimendu-analisiaren ikuspegitik, hitz multzoaren sentimendu-balentzian ez da aldataririk gertatzen. Izan ere, kasu honetan, ezeztapena ondotik informazioa gehitzeko erabiltzen da eta ez ezeztapenaren irismena den *ikuspuntu politikotik* ezeztatzeko. Beraz, *ez ezik* egiturak bere irismenaren parte diren hitzen sentimendu-balentzian ez du eragiten.

Erregeletan oinarritzen diren SO-CAL moduko (Taboada *et al.*, 2011) tresnetan ezeztapenari buruzko informazioa gehitu nahi bada, beharrezkoa da ezeztapen-markak eta haren irismena identifikatzea eta, horregatik, ezeztapen-markak eta beren irismena identifikatzeko erregelak sortu eta ebaluatu ditugu¹³, 3.18 taulan ikus daitekeen moduan.

3.18 taulako, erregelak deskribatu aurretik, erregela horiek nola sortu diren azalduko dugu jarraian, (38) eta (39) adibideen bitartez. Bertan bosgarren erregelako bi aldaerak landuko ditugu. (38) adibidean, *ezin* ezeztapen-markaren ondoren, aditz laguntzailea (*da*) eta aditz nagusia (*baztertu*) ageri dira eta, azkenik, izen-sintagma bat den subjektua (*ekaitz zaparradaren bat izatea*). Beraz, esaldi honetarako lortutako egitura honelakoa izango litzateke: *ezin* + aditz laguntzailea + aditz nagusia + IS. Dena kako artean dago eta horrek ezeztapen-markaren irismena zein den adierazten du.

(38) (...) [ezin da baztertu₋₁ ekaitz zaparradaren bat izatea.] (EGU35)

(39) [Irabazi₊₂ ezinik] jarraitzen du Eibarrek (...) (KIR22)

(39) adibidea aztertzen badugu, ikusiko dugu *ezin* ezeztapen-markaren aurretik *irabazi* aditz nagusia agertzen dela eta bera dela ezeztapen-markaren irismena. Corpusean ageri diren *ezin* ezeztapen-markaren beste 36 instantziekin prozedura bera jarraitu ondoren, 3.18 irudiko (1) eta (5) adibideetan azaltzen diren erregelak lortu ditugu.

Bildu den informazioetako bat erregela bakoitzak duen egitura sintaktikoa da. 3.18 taulako erregeletako egiturek honelako ezaugarriak di-

¹³Ezeztapen-markak eta beren irismena identifikatzeko *Murritzapen Gramatikan* (Karls-son *et al.*, 1995) sortutako eta ebaluatutako erregelei buruzko informazio gehiago 5.2.3 atalean dago.

Adb.	Ezeztapen-marka	Erregelen egitura
(1)	ezin	PM <i>ezin</i> + [adjektiboa/adberbioa] (+ atzizki konp.) PM
(5)	ezin	PM [(IS) + aditza + <i>ezin</i>] PM PM [<i>ezin</i> (+ ad. lag.) (+ IS) + aditza (+ IS)] PM
(10)	ez	PM [IS] <i>ez ezik</i> PM

3.18 taula: Sentimendu-balentzian eragin ezberdinak dituzten ezeztapenak identifikatzeko proposatu diren erregelako batzuk.

tuzte: kakoek [] ezeztapen-markaren irismena adierazten dute, parentesiek () sintagma horren agerpena hautazkoa dela adierazten dute. *Letra etzanak* ezeztapen-marka edo egitura lexikalizatuen osagaiak adierazten dituzte eta barrak /, azkenik, ezeztapen-markaren irismenean, aukeran aipaturiko osagai ezberdinek daudela adierazten du. Erregeletan ageri diren beste elementuak hitz-multzo eta gramatika-kategoriei dagozkie. ISk eta ASk izen- eta aditz-sintagma adierazten dute, hurrenez hurren. Erregeletan ageri diren beste elementuak dira adjektiboa, adberbioa, aditza, aditz laguntzailea eta sintagma. Bukatzeko, bada beste elementu bat erregeletan azaltzen dena eta hori puntuazio-markaren murriztapena (PM) da. Murriztapen honen helburua erregeletan ezeztapen-marken irismena esaldi barnean kokatzea eta beste esaldietako elementuak ez hartzea da. Izan ere, irismena ezeztapen-markaren aurrean nahiz atzean ager liteke eta aurreko edo ondorengo esaldietako hitzak ezeztapen-markaren irismeneko hitz moduan tratatzeko arriskua ikusten dugunez, erregelei puntuazio-marka murriztapena gehitzea erabaki dugu. Egitura lexikalizatuen kasuan, ez dago horrelako murriztapenik. Egitura lexikalizatuek egitura egonkorra dutenez (errepikatzen diren hitz-multzoak direnez) ez da beharrezkoa. Egitura lexikalizatuaren adibide bat “baino ez”. Bi hitzak elkarrekin uste baino maiztasun handiagoz azaltzen dira eta bi hitzen baturak berezko esanahi bat sortzen du, “bakarrik” hitzak duen esanahiaren parekoa dena.

Aurreko urratsetan analisi linguistikoan bildutako informazioa modu egitura-tuan antolatu ondoren (3.18 taulan ikusten den moduan), ezeztapen-markak eta beren irismena identifikatzeko erregelak Karlsson *et al.*en (1995) *Mu-*

rriztapen Gramatika hurbilpenean sortu eta ebaluatu ditugu. 3.7 irudian erregelak nolakoak diren ikus daiteke¹⁴. Erregela horiek honela osatuta daude:

```
LIST PUNTUAZIOA = PUNT.PUNT PUNT.KOMA PUNT.BI.PUNT PUNT.GALD PUNT.ESKL
PUNT.HIRU PUNT.PUNT.KOMA;

LIST EZ = ‘‘ez’’;
LIST BESTERIK = ‘‘beste’’;

# (2)
# besterik ez egiturak
MAP (!besterikezHAS) TARGET (DET) IF (0C BESTERIK) (1C EZ);

(...)
```

3.7 irudia: Ezeztapen-markak eta beren irismen-eremua identifikatzeko erregelen adibide bat *Murriztapen Gramatika* (Karlsson *et al.*, 1995) ingurunean.

- LIST zerrendak. Bertan hitz jakinak edo bestelako elementuak, puntuazio-markak adibidez, zerrendatzen dira. 3.7 irudian, puntuazio-markak (LIST_PUNTUAZIOA) eta ezeztapen-markak (LIST_EZ, LIST_BESTERIK) zerrendatu ditugu.
- MAP komandoa (islapen-erregelak). Hauen bidez erregeletan adierazitako egiturak corpusean bilatzen dira eta aurkitzen badira, etiketa bat esleitzen zaio corpusean bilatu den egiturari. Ezaugarri hauetaz osatuta dago erregela:
 - Esleituko den etiketa. Etiketak aurretik ! ikurra du, 3.7 irudian.
 - Etiketa esleituko zaion hitza. Adibidez, *!besterikezHAS* etiketa determinatzaile bati (DET) esleituko zaio.
 - Erregelaren baldintzak. Esaterako, *!besterikezHAS* etiketa esleitzeko, 0 posizioan BESTERIK zerrendak, hau da, *beste* hitzak, (0C BESTERIK) egon behar du eta 1 posizioan EZ zerrendak (kasu honetan, *ez* ezeztapen-markak), erregelak funtziona dezan.

```

“<, >” “<PUNT.KOMA>”
    PUNT.KOMA
“<batere >”
    “batere” ADB ARR ZERO w39,L-A-ADB-ARR-13,lsfi48 @ADLG %SINT
“<harrotu >” S:278/0
    “harrotu” ADI SIN PART NOIDEK w40,L-A-ADI-SIN-38,lsfi49 @-JADNAG
    %ADIKAT S:278 !gabeAUR
“<gabe >” S:141/0
    “gabe” ADB ARR ZERO w41,L-A-ADB-ARR-14,lsfi50 @KM▷ %SINT S:141 !
    gabe
“<$. >” “<PUNT.PUNT>”
    PUNT.PUNT

```

3.8 irudia: *Murritzapen Gramatikan* (Karlsson *et al.*, 1995) sortutako erregelek esleitutako ! etiketa.

Murritzapen Gramatikan (Karlsson *et al.*, 1995) idatzirik erregela bakoitzak bere etiketa uzten du corpuseko lerroan. 3.8 irudian, esaterako, *Murritzapen Gramatikan* sortu ditugun erregelek bi hitzi etiketa ezarri diete (!gabeAUR etiketa *harrotu* hitzean eta !gabe etiketa *gabe* hitzean).

Guk sorturiko erregelek corpusean ! ikurra duen etiketak esleitzen dituztela begiztatu ondoren, erregela horiek ebaluatu ditugu.

Ebaluazioa egiteko, Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) 48 testu erabili ditugu. 48 testu horiek prozesatu egin behar izan ditugu *Murritzapen Gramatikaren* ingurunera (Karlsson *et al.*, 1995) ekartzeko eta horretarako *Murritzapen Gramatikan* (Karlsson *et al.*, 1995) oinarritzen den euskarazko desanbiguatzaile morfosintaktikoa erabili dugu (Aduriz *et al.*, 1997).

Murritzapen Gramatikaren (Karlsson *et al.*, 1995) ingurunean sorturiko 75 erregelak corpusaren test zatiko 144 testuetan aplikatu ondoren, erregela horiek ebaluatu egin ditugu. Lehenik eta behin, bi pertsonen ebaluazioa egiteko duten adostasuna neurtu dugu corpusaren zati txiki bat erabiliz. Adostasuna neurtzeko kappa neurria (Cohen, 1960) erabili da. Emaitzek erakusten dutenez, kappa neurria 0,60koa da. Landis eta Kochen (1977) arabera, *neurritzko adostasuna* da. Etiketatzaileen arteko adostasuna neurtu ondoren, etiketatzaile batek ebaluazio osoa egin du.

¹⁴Ezeztapen-markak eta beren irismena identifikatzeko sortutako erregela guztiak A eranskinean daude zerrendatuta.

Ebaluatzeko, erregelek corpusean esleitutako etiketak ebaluatu ditu etiketatzaileak. Etiketatzaileak hiru etiketa erabili ditu: ETIK_ONDO, erregelaren etiketa zuzena denean; ETIK_FALTA, corpuseko hitzak etiketa behar zuzenean eta erregelak jarri ez dionean eta, azkenik, ETIK_GAIZKI, erregelak corpuseko hitzari etiketa jarri dionean eta hitzak etiketa behar ez duenean.

Ebaluazioaren adibide bat ikus daiteke, 3.9 irudian. Bertan hiru hitz daude eta hirurek egon beharko lukete etiketatuta. Hiruretatik bi (*harrotu* eta *gabe*) etiketatuta daude eta ETIK_ONDO etiketa jarri zaie, lerroen hasieretan. Baina, hori ez da hala *batere* hitzaren kasuan. Nahiz eta ezeztapen-markaren irismen-eremuko parte izan, *harrotu* aditzari eragiten diolako, ez dago etiketatuta. Ondorioz, hitz horri ETIK_FALTA etiketa jarri zaio, kasu honetan ere, lerroaren hasieran. Aipaturiko prozedura hau jarraitu da bi pertsonen arteko adostasuna nahiz corpusa ebaluatzerakoan.

```

“<, >” “<PUNT.KOMA>”
PUNT.KOMA
ETIK.FALTA “<batere >”
“batere” ADB ARR ZERO w39,L-A-ADB-ARR-13,lsfi48 @ADLG %SINT
ETIK.ONDO “<harrotu >” S:278/0
“harrotu” ADI SIN PART NOTDEK w40,L-A-ADI-SIN-38,lsfi49 @-JADNAG
%ADIKAT S:278 !gabeAUR
ETIK.ONDO “<gabe >” S:141/0
“gabe” ADB ARR ZERO w41,L-A-ADB-ARR-14,lsfi50 @KM> %SINT S:141 !
gabe
“<$.>” “<PUNT.PUNT>”
PUNT.PUNT

```

3.9 irudia: Erregelak ebaluatzeko etiketatzaileak hitzei esleitutako etiketak.

3.2.3. Diskurtsoa: erlazio diskurtso-egiturak, beren osagaiak eta unitate zentrala

Diskurtso mailan, RST jarraituz¹⁵ (Mann eta Thompson, 1988), bi alderdi landu ditugu. Alde batetik, balentzia-aldaketak erlaziozko diskurtso-egituretan aztertu ditugu. Beste aldetik, unitate zentraletan, hots, testuko

¹⁵*Egitura Erretorikoaren Teoriari* (RST) buruzko azalpenak 5.3.2 azpiatalean daude.

diskurtso-unitate garrantzitsuetan, orientazio semantikoa duten hitzek zer gramatika-kategoritakoak diren landu dugu.

3.2.3.1. Erlaziozko diskurtso-egiturak

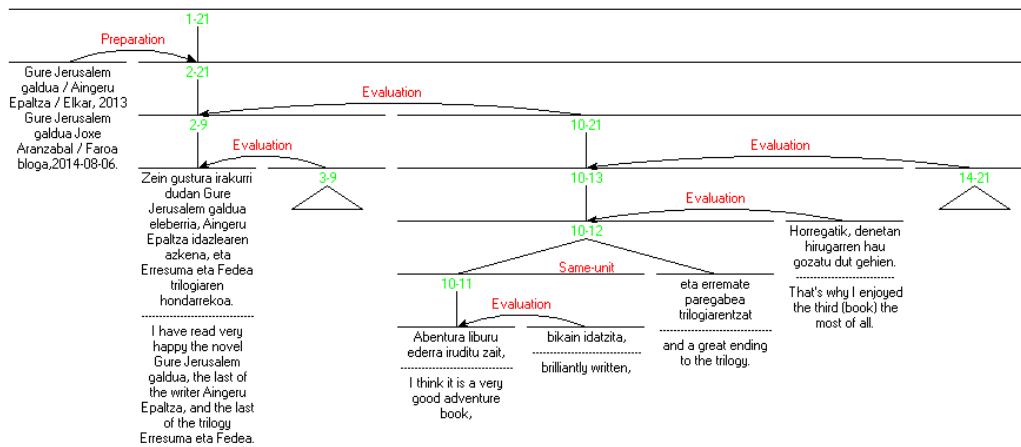
Erlazio erretorikoetan, lehenik eta behin, gure helburuak zeintzuk diren zehaztu ditugu. Guztira hiru helburu finkatu ditugu:

- Erlaziozko diskurtso-egituren eta bere osagaien arteko orientazio semantikoaren arteko adostasuna neurtu nahi dugu.
- Erlaziozko diskurtso-egiturek unitate zentraletik duten distantzia aintzat hartuz, erlazio erretorikoek eta testu osoak zer distantzian duten orientazio semantikoaren adostasunik handiena neurtu nahi dugu.
- Iritzi-testuetan, esanahi erretoriko ezberdineko erlazioek testuetan bere kokagune zehatza baduten aztertu nahi dugu. Erlazio erretorikoen kokagunea unitate zentralarekiko distantzian oinarrituz finkatuko dugu.

Erlaziozko diskurtso-egituretan ditugun helburuak finkatu ostean, zenbait urrats bete ditugu. Jarraian, metodologia urratsez urrats deskribatzen da:

- 1- Diskurtso mailako osagai ezberdinei orientazio semantikoa esleitu diegu.
 - 240 iritzi-testuri.
 - Erlaziozko diskurtso-egitura hauei (29 iritzi-testuetatik hartutakoak): EBALUAZIOA/INTERPRETAZIOA (140), KAUSA azpitaldea (71), KONTZETSIOA/ANTITESIA (70), EBIDENTZIA/JUSTIFIKAZIOA (53), KONTRASTEIA (26), BALDINTZA azpitaldea (18), AHALBIDERATZEA/MOTIBAZIOA (6). Guztira, 384 instantziei esleitu zaie orientazio semantikoa.
 - Aurreko erlazio erretorikoen osagaiei: nukleoa eta sateliteari edota erlazioko lehen eta azken osagaiari esleitu zaie orientazio semantikoa.

- 2- Erlaziozko diskurtso-egituretako parametroen analisisia egin dugu. Literaturaren domeinuko 28 testuen erlazio erretorikoak datu-base batean sartu ostean, jarraian zerrendatuta dauden parametroak aztertu ditugu. 3.10 irudiko EBALUAZIOA erlaziozko diskurtso-egitura (EDU 10-12) adibide moduan erabiliko dugu parametroak esplikatzeko:



3.10 irudia: SENTFAR-01 testuaren RST-zuhaitza.

- Nuklearitatea. Erlaziozko diskurtso-egitura nukleo bakarrekoa edo nukleo anitzekoa izan daiteke. 3.10 irudian, EBALUAZIOA erlaziozko diskurtso-egitura N(ukleo)-S(atelitea) motakoa dela zehaztu dugu.
- Erlaziozko diskurtso-egituren eta EDUen orientazio semantikoa. Erlaziozko diskurtso-egiturari nahiz EDUei hiru motako orientazio semantikoa esleitu diegu: positiboa, negatiboa eta neutrala. Lehenik eta behin, EDUei esleitu diegu orientazio semantikoa eta ondoren, erlazio erretoriko osoari. 3.10 irudiko EBALUAZIOA erlazioan (10-11 eta 12), bi EDUei eta erlazio osoari orientazio semantiko positiboa esleitu diegu.
- Unitate zentralarekiko distantzia. Erlaziozko diskurtso-egituren eta unitate zentralaren arteko distantzia bien artean dauden erlaziozko diskurtso-egituren kopurua zenbatuz neurtu dugu. Aztertzen

ari garen EBALUAZIOA erlazioaren kasuan (EDU 10-12), distantzia +2 dela adierazi dugu, unitate zentrala 3.10 irudian ezkerretik hasita bigarren EDUa baita eta tartean bi erlaziozko diskurtso-egitura baitaude.

- Erlaziozko diskurtso-egituren esanahi erretorikoa. Erlazioen esanahi erretorikoa izan da aztertutako azken parametroa. Adibidetzat hartu dugun erlazioa EBALUAZIOA motakoa da. Horrek 10-11 EDUek egoera bat aurkeztu eta 12 EDUak egoera horri buruzko ebaluaziozko aipamen bat egiten du.

- 3- Erlaziozko diskurtso-egituretan eta beren osagaiei orientazio semantiko etiketatzaileen artean dagoen adostasuna neurtu dugu. Eskuzko ebaluazioa egin dugu, F-neurria erabiliz.

3.2.3.2. Unitate zentrala

Bestetik, unitate zentrolean zer gramatika-kategoritako hitzak diren ohikoenak jakiteko eta orientazio semantikoa duten hitzek nolako banaketa duten aztertzeko, jarraian zerrendatuta dauden urratsak burutu ditugu:

- 1- Corpusetik unitate zentralak atera. Lehenik eta behin, pertsona batek Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) 192 testu hartu eta testu bakoitzaren unitate zentrala hautatu du. Test zatiko testuak ez dira erabili. Unitate zentrala hautatu aurretik, testuak segmentatu egin dira RST hurbilpeneko Das eta Taboadaren (2018) gidalerroei jarraituz. Bestalde, testuen segmentazioa egin ahal izateko, *RSTTools* (O'Donnell, 2000) programa erabili dugu.

3.11 irudian, EGU01 testua segmentatuta ikus daiteke XML formatuan. Segmentu horietako bakoitza oinarrizko diskurtso-unitate bat da eta anotatzailearen zeregina, hurrengo urratsarekin lotura eginez, diskurtso-unitate horietako zein den unitate zentrala, hots, iritzi-testuko diskurtso-unitaterik garrantzitsuena, identifikatzea da.

- 2- Unitate zentralaren hautaketako adostasuna neurtu. Pertsona horrek unitate zentrala ondo hautatu duen frogatzeko, beste pertsona batek

```

<rst>
  <header>
    <relations>
    </relations>
  </header>
  <body>
    <segment id="1">Gaurko eguraldia. HEGO EKIALDEKO HAIZEA ETA HODEI
      BATZUREKIN EPELDU EGINGO DA.</segment>
    <segment id="2"> Giro ona tokatuko zaigu gaur ere hego haize
      epelarekin.</segment>
    <segment id="3"> Borraska pixka bat hurbiltzeak haizea apur bat
      mugituko du</segment>
    <segment id="4"> baina </segment>
    <segment id="5">tenperatura igoaz , </segment>
    <segment id="6">giroa goxo mantenduko da leku gehienetan.</segment>
    <segment id="7"> Hodei zirrinta mehe batzuk agertuko dira zero sabaian
      </segment>
    <segment id="8"> baina eguzkia apur bat lausotu arren,</segment>
    <segment id="9"> astro handia bistan edukiko dugu</segment>
    <segment id="10"> eta dotoreziak ez du bat ere galduko.</segment>
    <segment id="11"> Baditeke iluntze aldera ekaitz hodei batzuk garatzea
      eta euri zarrastaren batzuk botatzea agian. </segment>
    <segment id="12">Akats txiki horiek gora behera eguraldi bikaina oro
      har.</segment>
  </body>
</rst>

```

3.11 irudia: Bi pertsonen egin beharreko unitate zentralaren aukeraketa EGU01 iritzi-testuan.

ere testu horien unitate zentralak hautatu ditu. Bi pertsonen arteko adostasuna neurtzeko ez dira unitate zentral guztiak erabili; izan ere, guztira 78 iritzi-testuren (corpus osoaren % 32,5) unitate zentrala hautatu behar izan dute bi pertsonen. Adostasuna kalkulatzeko F-neurria erabili dugu.

Adostasunaren neurketak 3.19 taulan erakusten dira, 78 iritzi-testuetatik 51etan, bi pertsonen diskurtso-unitate bera jo dute unitate zentraltzat. Beste 44 iritzi-testuetan, aldiz, diskurtso-unitate ezberdinak lotu dituzte unitate zentralarekin. Beraz, 78 iritzi-testuren unitate zentrala aukeratzean egon den adostasuna 0,65ekoa izan da.

- 3- Unitate zentraletako hitzen gramatika-kategoriaren azterketa. Hurrengo urratsean, unitate zentral bakoitzean zer hitz-mota agertzen den aztertu nahi izan dugu, baita hitzen bat edo batzuk maiztasun erre-

Neurketa	Instantziak	F-neurria
Adostasuna	51	0,65
Desadostasuna	27	0,35

3.19 taula: Bi pertsonen arteko adostasun-maila 78 iritzi-testuetan unitate zentrala aukeratzekoan.

gular batez azaltzen diren ere. Horretarako, *Analhitza* tresna (Otegi *et al.*, 2017) erabili dugu. Aipaturiko *Analhitza* tresna horretara, hautatuak izan diren unitate zentral guztiak igo ditugu. Unitate zentral horiek domeinuka aztertu ditugu tresnaren bidez. Unitate zentralak aztertzeko, honako alderdiak hartu ditugu kontuan:

- i) Hitzen maiztasuna. Gramatika-kategoria bakoitzeko, gehienez maiztasun handieneko 30 hitz aztertu ditugu.
- ii) Hitzen-gramatika kategoria. Aintzat hartutako hitzen besten ezau-garri bat beren gramatika-kategoria izan da. *Analhitzak* (Otegi *et al.*, 2017) zuzenean hitzak gramatika-kategoria aintzat hartuz sailkatzen ditu; horrenbestez, lana erraztu digu.

3.20 taulan, *Analhitza* tresnak Otegi *et al.* (2017) eguraldiaren domeinuko hitzak zein gramatika-kategoriatan sailkatu dituen ikus daiteke, Izen-kategoriako hitzak dira gehienak (% 32,65).

Gramatika-kategoria	Kopurua	%
Izenak	80	32,65
Adjektiboak	25	10,20
Aditzak	69	28,16
Adberbioak	19	7,76
Determinatzaileak	10	4,08
Juntagailuak	35	14,29
Preposizioak	0	0,00

3.20 taula: *Analhitza* tresnak (Otegi *et al.*, 2017) eguraldiaren domeinuko hitzei egindako gramatika-kategoriaren sailkapena.

Gainera, *Analhitza* tresnak (Otegi *et al.*, 2017) gramatika-kategoria

bakoitzeko hitzak zerrendatuta ematen ditu, 3.21 taulan ikusi daitekeen moduan.

Adjektiboak
epel
eder
dotore
goxo
eguzkitsu
atsegin
txar
oker

3.21 taula: Eguraldiaren domeinuko unitate zentralerako hitzen zerrenda.

- iii) Hitzen domeinua. Maiztasunik handieneko hitz horiek domeinuari oso lotutakoak diren edo ez ere aztertu dugu. Domeinu bakoitzean, hitzak bi multzotan sailkatu ditugu: hitzak semantikoki domeinukoak diren edo ez. Eguraldikoaren kasuan, hitzak hiru multzotan sailkatu ditugu: i) eguraldi domeinuari loturiko hitza, ii) eguraldiaren domeinuarekin zerikusirik ez duen, iii) denboradierazpen bat den. Eguraldia domeinuan, denborak garrantzia duela uste dugulako egin ditugu hitzen sailkapenean hiru multzo. Esaterako, eguraldiaren domeinuko maiztasunik handieneko hitzetan, *negu* hitza domeinuko moduan sailkatu dugu eta *itxura*, berriz, ez; ez dagoelako zuzenean domeinu horri lotuta. Azkenik, *asteburu* hitzak denbora adierazten duenez, denborazkoetan sailkatu dugu.
- 4- Unitate zentralerako eta horien hitzerako sentimenduen balentzia esleipena. Aztertu dugun beste ezaugarrietako bat unitate zentralerako gramatikakategoriaren sentimendu balentzia izan da. Sentimendu-balentziaren esleipena egiteko sortu dugun euskarazko sentimenduen sailkatzailea erabili dugu.

(40) adibidean, euskarazko sentimenduen sailkatzaileak eguraldiaren domeinuko unitate zentral bati sentimendu-balentzia nola esleitu dion ikus daiteke. *Ezegonkor* adjektiboari esleitu dio sentimendu-balentzia eta gramatika-kategoria bakoitzean zenbat hitzei esleitu zaien sentimendu-balentzia zenbatu dugu.

(40) giro [*ezegonkor*]_{-1,5} nagusi izan izan etorri egun .

3.22 taulan, eguraldiaren domeinuko gramatika-kategorietan, zenbat hitzek duten sentimendu-balentzia zenbatzen da. Adjektibo-kategoriako hitzen dute sentimendu-balentzia gehien.

EGURALDIA	Sentimendu-balentziadun hitzen kopurua
Izena	7
Aditza	12
Adjektiboa	22
Adberbioa	3
Guztira	44

3.22 taula: Eguraldiaren domeinuan, gramatika-kategoria bakoitzeko sentimendu-balentziadun hitzen kopurua.

3.3 Laburpena

Kapitulu honetan, tesi-lan honen metodologia deskribatu dugu. Lehenik eta behin, balentzia-aldatzaileak identifikatzeko eta beren eragina neurtzeko baliabide eta tresnak sortu ditugu: Euskarazko Iritzi Corpora (Alkorta *et al.*, 2016) (diskurtso-egituraren eta orientazio semantikoaren aldetik etiketatuta) eta *SentiTegi* (Alkorta *et al.*, 2018) izeneko euskarazko sentimendu lexikoa.

Bigarren urratsean, balentzia-aldatzaileak identifikatu eta euren eragina neurtu dugu. Balentzia-aldatzaile fonologikoak eta morfologikoen instantziak corpus zati batetik eta sentimendu-balentzia esleitu diegu. Gero, instantzia horiek hitzean eragiten duten neurtu dugu.

Ezeztapen-markak ere antzeko modu batean landu ditugu. Ezeztapen-markak corpus zati batean bilatu eta ezeztapen-marka duten esaldiei sentimendu-balentzia esleitu diegu. Beren eragina neurtu ondoren, ezeztapen-markak eta beren irismena identifikatzeko erregelak garatu eta, azkenik, *Murritzapen Gramatika* ingurunera egokitu ditugu horiek ebaluatzeko asmoz.

Diskurtso-mailan, RST hurbilpenaz etiketatutako corpus zatitik esanahi erretoriko ezberdinetako erlazio lortu ditugu. Ondoren, eskuz nahiz *SentiTegi* (Alkorta *et al.*, 2018) erabiliz orientazio semantikoa eta sentimendu balentzia esleitu diegu eta erlazio erretorikoak sentimenduen analisiaren ikuspegitik landu ditugu. Bestalde, unitate zentrala ere landu dugu.

Sentimenduen analisirako baliabideak

Kapitulu honetan, tesi-lan honen ondorioz garaturiko baliabideak aurkeztuko ditugu. Guztira hiru baliabide garatu ditugu: Euskarazko Iritzi Corpusa (4.1 atala), *Sentitegi* izeneko euskarazko sentimenduen lexikoa (4.2 atala) eta, azkenik, dokumentu mailako euskarazko sentimenduen sailkatzailea (4.3 atala).

4.1 Euskarazko Iritzi Corpusa

Atal honetan, euskarazko egunkari eta webgune espezializatuetik bildutako 240 iritzi-testuez osaturiko corpusa azalduko dugu. Lehenik eta behin, 3.1.1 atalean jarraituriko metodologiaz sortu dugun iritzi-testuen corpusaren ezaugarriak aipatuko ditugu. Ondoren, berriz, metodologiaren garapenean izandako zailtasunen berri emango dugu.

4.1.1. Euskarazko Iritzi Corpusaren ezaugarriak

Sortu dugun Euskarazko Iritzi Corpusak guztira 240 iritzi-testu ditu sei domeinutakoak: kirola, politika, musika, zinema, liburuak eta eguraldia.

Corpusaren beste ezaugarrietako bat bertako iritzi-testuetan agertzen den balorazioaren oreka da. Domeinuko 40 testu daude eta erdiek balorazio positiboa dute eta beste hainbestek, berriz, negatiboa.

Analhitza tresnak (Otegi *et al.*, 2017) ematen dituen emaitzen arabera, corpusak 52.092 token eta 3.711 esaldi ditu. Corpusa beste hizkuntzetako iritzi-

testuen corpusekin alderatuz, tamainaz zertxobait txikiagoa dela esan daiteke, 4.1 taulan agertzen den moduan. Beste hizkuntzetan, oro har, iritzi-testu kopurua handiagoa da, baina, tokenen kasuan, euskarazko corpusaren eta beste corpusen arteko ezberdintasuna are handiagoa da.

Corpusa	Testuak	Hizkuntza	Tokenak
Boldrini <i>et al.</i> (2010)	300		
Hu eta Liu (2004)	113		81.855
Rushdi-Saleh <i>et al.</i> (2011)	500	Arabiera	215.948
Wilson (2008)	535	Ingelesa	265.000
Taboada (2008)	400	Ingelesa	289.270
Quan eta Ren (2009)	1.487	Txinera	878.164

4.1 taula: Sentimenduen analisirako eskura dauden iritzi-testuen corpusak.

4.1.2. Euskarazko Iritzi Corpusaren garapena

Euskarazko Iritzi Corpusa sortzerakoan, bi zailtasun izan ditugu. Alde batetik, guk nahi genituen moduko iritzi-testuak bilatzea zaila suertatu da (4.1.2.1 atala). Beste aldetik, domeinu batzuetan, orientazio semantiko jakin batzuetako iritzi-testuak biltzea ere asko kostatu da (4.1.2.2 atala). Azkenik, 4.1.3.2 atalean, corpusaren baliagarritasuna neurtzeko hautatutako irizpi-deak zergatik hautatu ditugun azalduko dugu.

4.1.2.1. Ezaugarri jakin batzuetako iritzi-testuak bilatzeko zailtasunak


Gure lana eta lanaren emaitza konparatu ahal izateko, badagoen baliabide (kasu honetan, corpusa) baten antzeko bat egin nahi izan dugu. Hau da, hitz edo sintagmen balentzia aldatzen duten fenomeno linguistikoak aztertzeko moduko iritzi-testuak bildu nahi ditugu. Horregatik, Euskarazko Iritzi Corpusa sortzerakoan, *SFU Review Corpus*¹ (Taboada, 2008) hartu dugu erreferentziatzat.

¹Corpusa eskuragarri dago https://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html webgunean.

SFU Review Corpusak (Taboada, 2008) zortzi gaietako 400 iritzi-testu biltzen ditu. Gai bakoitzeko 50 testu daude, erdiak iritzi positibodunak dira eta beste erdiak, berriz, negatibodunak. Iritzi-testuek lantzen dituzten gaiak hauek dira: liburuak, autoak, ordenagailuak, sukaldeko tresneria, pelikulak, musikak eta telefonoak.

Gure asmoa gai horiei buruzko iritzi-testuak biltzea izan da, baina euskara hizkuntza gutxitua denez, eta horrek zailtasunak areagotu egiten dituenaz, asmoa birplanteatu egin behar izan dugu. Izan ere, liburu, musika eta pelikulei buruzko iritzi-testuak aurkitu ditugu, baina besteen kasuan, kalitate baxukoak edo desagokiak diren iritzi-testuak aurkitu ditugu edo ez ditugu aurkitu.

4.1 irudian, *Tripadvisor* webgunetik hartutako iritzi-testu bat ikus daiteke. Hiru hizkuntzetan idatzita dago iritzi-testu bakarra eta hiruretan esaten dena ez da guztiz berdina. Hizkuntza bat baino gehiago agertzen diren iritzi-testuak aurkitzea oso ohikoa izan da gure iritzi-testuen bilaketan. Badirudi, iritzi-emaile euskaldunek beren iritziak ahalik eta irismen handiena izatea nahi dutela eta, horregatik, euskaraz gain beste hizkuntza bat erabiltzen dutela.




 Opinión escrita el 11 de marzo de 2014




Nice hotel

Nice hotel in Dublin centre close to Christchurch. I, Il repeat when i, Il come back to Dublin
Hotel agradable en el centro de Dublin. El personal algo seco, y eso que los irlandeses presumen o dice que son hospitalarios, pero bien, algo caro pero es la tónica habitual en Dublin
Oro har, atsegina baina garesti samar kontutan zaharra dela baina oso ondo kokatuta,

Fecha de la estancia: marzo de 2014

Tipo de viaje: Viajé con mi familia

 Relación calidad-precio
 Ubicación
 Calidad del sueño

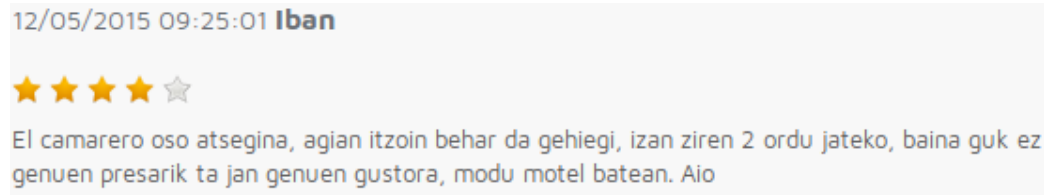
 Habitaciones
 Limpieza
 Servicio

[Pregunta a mikelasbilbo sobre Harding Hotel](#)

4.1 irudia: Iritzi-testu hirueledun baten adibidea.

Halaber, aurkitu ditugun euskarazko iritzi-testu gehienak labur samarrak dira eta horrek gure ikerketa-lana mugatu dezake, aztertu nahi ditugun hizkun-

tza-gertakariak maiztasun txikiagoz agertu direlako. Euskarazko iritzi-testu labur bat ageri da 4.2 irudian.



4.2 irudia: Euskarazko iritzi-testu labur baten adibidea.

Ikusten denez, ortografia zuzena ez izateaz gain, egitura sintaktikoa ez da aberatsa eta bertatik diskurtsoaren egitura lantzea ere zailagoa bilakatzen da, testuaren tamaina ez delako handia. Egoera horren ondorioz, zenbait aldizkari eta komunikabide aztertu ditugu eta aurkitzen errazagoak diren beste gai batzuk dituzten iritzi-testuak biltzea pentsatu dugu. Hala, liburuei, musikari eta pelikulei buruzko iritziak biltzeaz gain, eguraldiari, politikari eta kirolari buruzko iritzi-testuak biltzea erabaki dugu. Gainera, ahalik eta kalitate hobereneko iritzi-testuak biltzeko, komunikabide eta aldizkari espezializatueta jotzea erabaki dugu.

Nahiz eta iritzi-testuen gaiak batzuetan *SFU Review Corpusekoetatik* (Taboada, 2008) urrundu diren, beste ezaugarrietan antzekoak izan daitezten saiatu gara. Horregatik, erabaki dugu Euskarazko Iritzi Corpusak domeinu bakoitzeko 40 testu izatea eta domeinu bakoitzeko, iritzi-testuen erdiak positiboak izatea eta besteak negatiboak izatea. Beraz, gai bakoitzeko 20 testu positibo eta beste hainbeste negatibo bildu ditugu.

4.1.2.2. Orientazio semantiko jakin bateko iritzi-testuak bilatzeko zailtasunak

Iritzi-testuak balorazioen aldetik orekatuak izateak beste zailtasun eta arazo batzuk ekarri dizkigu. Esaterako, literatura-liburuen kasuan, zailtasunak izan ditugu balorazio txarreko iritziak aurkitzeko. Zailtasun horren jatorriaren berri Egañak (2013) ematen du.

Egañak (2013) 1975tik 2005ra bitarteko 2.300 euskal literatura kritika aztertu ditu eta haren ikerketaren arabera, literatura kritiken % 84k iruzkin positiboa

edo oso positiboa dute. *Berria* egunkarian egindako elkarrizketan dioenez², “euskal literaturak berak ere bere burua legitimatzeko arazoak” ditu eta horren ondorio da iruzkin positiboen kopuru handia. Halaber, “Euskal literaturan oso jende gutxi da kritikari profesionala” eta “subjektuak eta legitimatua den objektuak elkarrengandik urrun samar” ez daudenez, legitimatzeko arazoak sortzen dira. Hots, euskal literatura komunitate txikia denez, eta denen arteko loturak beste literatura-komunitateekin alderatuta estuagoak direnez, iruzkin negatiboak gutxiago egiten dira. Egoera hori aintzat hartuta, balorazio negatiboak bereziki atzerriko lanei buruzko iritzi-testuetan aurkitu ditugu. Politikako iritzi-testuetan, berriz, aurkakoa gertatu da. 2008tik aurrera dagoen krisi ekonomikoa tarteko, iritzi gehienak iruzkin negatibokoak izan da eta iruzkin positibokoak aurkitzea zaila gertatu zaigu.

4.1.3. Euskarazko Iritzi Corpusaren baliagarritasunaren neurketa

Euskarazko Iritzi Corpusa osatu ondoren, corpus horrek subjektibotasunik baduen eta ezeztapen-marken instantzia ugari badauden eta, horrenbestez, sentimendu-analisirako baliagarria den neurtu dugu. Halaber, gure zenbait ikerketa-lerrotarako (hizkuntza maila ezberdinetako balentzia-aldatzaileak identifikatzeko eta euskarazko sentimenduen lexikoi bat sortzeko) beharrezkoak diren hizkuntza-fenomenoak ere agertzen ote diren aztertu dugu. Aztertzea erabaki ditugun aspektuak (lehen pertsonaren erabilera, adjektiboen agerpena eta ezeztapena) iturri bibliografikoetan oinarrituta hautatu ditugu. Jarraian, i) corpusaren baliagarritasuna neurtzeko erabili ditugun baliabideak eta irizpideak eta ii) neurketen emaitzak azalduko ditugu.

4.1.3.1. Baliagarritasuna neurtzeko baliabideak eta irizpideak

Euskarazko Iritzi Corpusaren subjektibotasuna neurteko lehen pertsonaren erabilera eta adjektiboen agerpena neurtu dugu eta, horretarako baliabideak 4.2 taulako corpusak izan dira. Corpus horiek alderatu egin ditugu.

²Elkarrizketa esteka honetan dago eskuragarri: www.berria.eus/paperekoa/1744/024/001/2015-02-04/literatura_batek_sendotzeko_gatazka_behar_du_bere_barruan.htm, *Berria*, 2015-02-14.

	Subjektiboa	Objektiboa
Euskara	Euskarazko Iritzi Corpusa	Euskarazko Wikipediako artikuluak
Ingelesa	SFU Review Corpus	Ingeleseko Wikipediako artikuluak

4.2 taula: Lehen pertsonaren eta adjektiboaren agerpena neurtzeko erabilitako corpusak.

- *SFU Review Corpus* (Taboada, 2008). Gure erreferentziazko corpusa da. Biek hainbat domeinutako iritzi-testuak biltzen dituzte eta, horrenbestez, eduki subjektiboa dute. 16.705 esaldi ditu corpusak.
- Ingeleseko Wikipediako zenbait artikulu. *SFU Review Corpus*aren (Taboada, 2008) domeinuekin lotura duten artikuluak hartu ditugu. Guztira 8 artikulu bildu ditugu eta corpus horrek 64.258 hitz ditu.
- Euskarazko Wikipediako zenbait testu. Ingeleseko Wikipedian bezala, kasu horretan ere, Euskarazko Iritzi Corpuseko domeinuekin bat datozen artikuluak erabili ditugu. Euskarazko Wikipediaren kasuan, 28 artikulu bildu ditugu eta corpus horrek 37.964 hitz ditu.

Wikipedia entziklopedia bat izanik, artikuluak ikuspegi neutroa edukitzea eskatzen da³. Testurik objektiboan horiek direla uste dugulako, ingelesezko eta euskarazko corpus objektiboak Wikipedian oinarrituta sortu ditugu.

Lehen pertsonaren eta adjektiboaren agerpenaren neurketa aipaturiko corpusetan horrela eta arrazoi horiengatik gauzatu dugu:

- 1- Lehen pertsonaren agerpena. Iritziak, normalean, lehen pertsonan adierazi ohi dira (Halliday *et al.*, 2014) lehen pertsona zenbateko maiztasunez agertzen den aztertu dugu gure corpusean. Gure corpuseko emaitzak konparatzeko 4.2 taulako beste hiru corpusak erabili ditugu.

Lehen pertsonaren agerpena neurtzerakoan, zailtasun batekin egin dugu topo. Euskaraz, aditz laguntzailan agertzen dira aditzari lotutako marka morfologikoak eta, horrenbestez, baita pertsona eta nume-

³Esteka honetan azaltzen dira Wikipediaren ezaugarriak: <https://eu.wikipedia.org/wiki/Wikipedia>.

roari buruzko informazioa ere. Ingelesean, aldiz, hirugarren pertsonaren kasuan izan ezik, aditz-forma guztiak berdinak dira eta horrek lehen pertsona neurtzea eragozten du. Konponbide moduan, ingelesean pertsona-izenordainen presentzia neurtzea erabaki dugu. Beraz, lehen pertsonaren agerpena neurtzeko, euskaran aditz laguntzaileak erabili ditugu eta ingelesean, ordea, pertsona-izenordainak. Neurketa egitean, lehen pertsona hori singularra edo plurala den kontuan hartu dugu.

- 2- Adjektiboen agerpena. (Liu, 2010, 32 orr.) lana gramatika-kategoriez (*Part-Of-Speech*, POS) ari denean, adjektiboak iritzien adierazle garrantzitsuak direla adierazten du. Horregatik, adjektiboek gramatika-kategoria guztien artean duten presentzia neurtu dugu.

Euskarazko Iritzi Corpusean ezeztapen-marka asko ageri diren zenbatu dugu; izan ere, ezeztapena nolako balentzia-aldatzailea den landutako hizkuntza-fenomenoetako bat izan da. Kasu horretan, bere agerpen asko dauden interesatzen zaigu eta ez beste corpusekin alderatzea; izan ere, ezeztapena lehen pertsona eta adjektiboen erabilera ez bezala ez dira subjektibitatearen adierazle.

- 3- Ezeztapena. Ezeztapena garrantzitsua da esaldien eta testuen sentimendu-balentzia aldatzeari dagokionez, Polanyi eta Zaenenen arabera (2006) testuinguruko balentzia-aldatzailea da. Hori dela eta, corpusean agertzen diren ezeztapen-markak zenbatzea erabaki dugu; instantzia kopuru minimo bat behar baitugu azterketa ondo gauzatzeko.

Laburbilduz, lehen pertsonaren eta adjektiboen agerpen kopurua beste corpusetako (euskarazko eta ingelesezko Wikipediako corpusak eta *SFU Review Corpus* (Taboada, 2008)) agerpen kopuruarekin alderatu dugu. Agerpen kopuruaren alderaketa eginez Euskarazko Iritzi Corpora tesi-lanerako baliagarria zaigun zehaztu dugu. Ezeztapenaren kasuan, aldiz, beren agerpena Euskarazko Iritzi Corpusean zenbatu egin dugu, aztertu dugun balentzia-aldatzaileetako bat baita.

4.1.3.2. Neurketen emaitzak

Corpusaren baliagarritasuna neurtzeko aztertu diren ezaugarriak hauek izan dira: lehen pertsonaren agerpena (neurketa eta alderaketa), adjektiboaren agerpena (neurketa eta alderaketa) eta ezeztapena (kontaketa).

- Lehen pertsonaren agerpena. Corpus horietan agertzen den lehen pertsona kopurua 4.2 taulako baliabideetakoekin alderatu dugu. Euskaran, lehen pertsonaren erabilera (singularra eta plurala) aditz laguntzailean neurtu dugu, eta ingelesean, pertsona-izenordainetan.

Corpusa	Lehen sg.	Lehen pl.	Guztira
Euskarazko Wikipedia	% 0,12	% 1,09	% 1,21
Ingeleseko Wikipedia	% 0,03	% 0,09	% 0,12
Euskarazko Iritzi Corpusa	% 3,27	% 5,10	% 8,37
<i>SFU Review Corpus</i>	% 10,10	% 1,70	% 11,80

4.3 taula: Lehen pertsonaren agerpena lau corpus ezberdinetan.

4.3 taulan agertzen den moduan, euskarazko nahiz ingelesezko Wikipediatatik hartutako testuekin osatutako corpusetan, lehen pertsonaren agerpena oso baxua da. Ingeleseko Wikipediaren kasuan, azaltzen diren pertsona-izenordain guztien % 0,12 lehen pertsonari dagokio. Joe-
ra bera erakusten du euskarazko Wikipediak, non bertako aditzetako marka morfologikoetan, % 1,21a lehen pertsonakoak diren.

Corpus subjektiboetan, aldiz, lehen pertsonaren agerpena handiagoa da. Euskarazko Iritzi Corpusean, aditzetako marka morfologikoetan, lehen pertsonaren agerpena % 8,37koa da. *SFU Review Corpusean* (Taboada, 2008), lehen pertsonaren agerpena handiagoa da; pertsona-izenordainen % 11,80 baitago lehen pertsonan. Bestalde, aipagarria da corpus subjektiboetan, lehen pertsona plurala dela nagusi euskararen kasuan eta lehen pertsona singularra, berriz, ingelesaren kasuan.

- Adjektibo-kopurua. Corpusaren baliagarritasuna neurtzeko aintzat hartu dugun beste alderdi bat adjektiboen kopurua izan da. Adjektiboak oinarrizko elementuak dira maila lexikoan subjektibotasuna erakusteko, beraz, corpus subjektiboetan adjektiboen agerpen handiagoa

espero liteke corpus objektiboetan baino. Hala ere, 4.4 taulako emaitzek ez dute hori erakusten.

Corpusa	Guztira
Euskarazko Wikipedia	% 8,50
Ingeleseko Wikipedia	% 9,09
Euskarazko Iritzi Corpusa	% 9,82
<i>SFU Review Corpus</i>	% 8,35

4.4 taula: Adjektiboen agerpena corpus objektibo eta subjektiboetan.

Emaitzen arabera, adjektiboen agerpena antzekoa da corpus guztietan, bai objektiboa edo subjektiboa izanda, baita euskarazkoa edo ingelesezkoa izanda ere. *Analhitzak* (Otegi *et al.*, 2017) emandako datuen arabera, Euskarazko Wikipedian eta ingelesezko Wikipedian, adjektiboak gramatika-kategoria guztietatik % 8,50 eta % 9,09 dira, hurrenez hurren. Euskarazko Iritzi Corpusean adjektiboak gramatika-kategoria guztien % 9,82 dira eta *SFU Review Corpusean* (Taboada, 2008) % 8,35. Ondorioz, badirudi adjektiboen agerpen handiagoa ez dela corpus subjektiboen ezaugarri bat. Baliteke horren ordez, adjektibo-mota izatea testu objektibo eta subjektiboen arteko ezberdintasuna.

- Ezeztapena. Corpusean neurtu dugun azken ezaugarria ezeztapena eta zehazki, ezeztapen-markak izan dira. Hiru ezeztapen-marka moten kontaketa egin dugu.

Alde batetik, perpaus osoari (adibidez, *gaur dendak ez daude zabalik*) eragiten dion *ez* ezeztapen-marka 718 aldiz agertzen da. Izen sintagmari (adibidez, *arrakastarik gabe*) eragiten dino *gabe* ezeztapen-marka 107 aldiz agertzen da eta azkenik, menpeko perpausari eragiten dien *ezean* lau aldiz zenbatu dugu. Kasu horietan, ezeztapen-marka kopuru handia aurkitu dugu corpusean eta horrek horiek ikertzeko aukera eman digu.

4.2 Euskarazko sentimenduen lexikoa

Atal honetan, hasteko, euskarazko sentimenduen lexikoiaren bi bertsioek dituzten ezaugarriak aurkeztuko ditugu. Ondoren, euskarazko lexikoiaren garapenerako zenbait alderdi (ordainen kasuistika eta hartutako erabakiak, besteak beste) aipatuko ditugu eta, azkenik, lexikoiaren ebaluazioa azalduko dugu.

4.2.1. Euskarazko sentimenduen lexikoiaren ezaugarriak

Gaztelaniazko SO-CAL lexikoa (Brooke *et al.*, 2009) itzultzeko jarraitu dugun prozesuaren ondorioz, euskarazko sentimenduen lexikoi beraren bi bertsio lortu ditugu. 4.5 taulak sentimenduen lexikoiaren bi bertsioen emaitzak erakusten ditu.

Gramatika-kategoria	V1.0		V2.0	
	Sarrerak	%	Sarrerak	%
Izenak	2.282	28,06	461	37,27
Adjektiboak	3.162	38,85	446	36,05
Aditzondoak	652	7,98	54	4,36
Aditzak	1.657	20,36	276	22,32
Intentsifikatzaileak	387	4,75		
Guztira	8.140	100	1.237	100

4.5 taula: Euskarazko sentimenduen lexikoiaren bi bertsioen ezaugarriak.

Sentimenduen lexikoiaren lehen bertsioa (V1.0) bigarren bertsioa (V2.0) baino handiagoa da. Horren atzean zenbait arrazoi daude. Hasteko, lehen bertsioan *Zehazki* (Sarasola, 2005) eta *Elhuyar* (Elhuyar, 2013) hiztegien bidez lorturiko euskarazko ordain guztiak hartu dira kontuan. Bigarren bertsioan, aldiz, bi hiztegi horietako sarrerak diren euskarazko ordainak bakarrik hartu dira kontuan. Horrenbestez, itzulpen bidez lortu diren lexikoiaren sarrerek aniztasun handia erakusten dute: batzuek izenlagun hizkia daramate (adibidez, *garrantzizko*), beste batzuk kolokazioak⁴ dira, etab. Halaber, intentsifi-

⁴Kolokazioak uste baino maiztasun handiagoaz batera agertzen diren hitzak dira. *Adarra jo* horren adibidea da.

katzaileak ere lexikoia lehenean badaude, bigarrenengoan ez bezala.

Lexikoi horien arteko berdintasunak eta ezberdintasunak aintzat hartuta, lexikoia lehenean 8.140 sarrera ditu, bigarrenengoan 1.237 dituen bitartean. Batetik bestera asko jaisten da sarrera kopurua; izan ere, bigarrenengo bertsioa domeinuari lotuta dago. Bi bertsioetan izenak eta adjektiboak dira gramatika-kategoria nagusienak. Aditzak zerbait gutxiago dira eta aditzondoak, berriz, are gutxiago. Bigarren bertsioan, intentsifikatzailek ez dago arrazoi praktikoengatik. Gure ustez, intentsifikatzaileek beste hitzei eragiten dietenez, gertuago daude sintaxi mailatik lexiko mailatik baino. Gainera, kontuan hartu behar da intentsifikatzaileek % -100 eta % +50 arteko balioa dutela biderketa eragiketean, beste gramatika-kategoriek -5 eta +5 arteko balioak (gehiketa eta kenketa eragiketak) dituzten bitartean.

Lexikoia bi bertsioen beste ezaugarri bat lexikoi paraleloak direla da. Hots, lexikoiaren sarrerekin datu-base bat sortuta dago eta euskarazko sarrerekin gaztelaniazko eta ingelesezko ordainak (bigarrenengo bertsioan) paraleloki eta modu ordenatuan daude jarrita. Gaztelaniazko eta ingelesezko ordainek ere beren sentimendu-balentzia badute, 4.6 taulan ikus daitekeenez. Bertan, lexikoia paraleloa dela erakusten duten lau adibide daude. Adibideak adjektiboak dira eta beren sentimendu-balentziak eskalan daude.

Hitza lexikoian	Balentzia	SPA	Balentzia	ENG	Balentzia
bikain	+5	excepcional	+5	excellent	+5
on	+2	buen	+2	-	-
eskas	-1	escaso	-2	insufficient	-1
txar	-3	adverso	-3	bad	-3

4.6 taula: Lexikoi paraleloaren adibide batzuk.

Ikusten den moduan, batzuetan zerrendetako sentimendu-balentziak ez datoz bat (*eskas* (-1), *escaso* (-2) eta *insufficient* (-1), esaterako), gaztelaniazko eta ingelesezko lexikoia ez direlako modu bere-berean sortu. Baina, euskarazko lexikoia sarreretako sentimendu-balentziak beti ingeleseko edota gaztelaniazko sarreretako sentimendu-balentzia batekin bat egin behar du.

Behin euskarazko lexikoia euskarazko lehenean bertsioa euskarazko sentimenduen sailkatzailean integratu ondoren, sorturiko euskarazko sentimen-

duen lexikoia gauza da hitzei nahiz esaldiei sentimendu-balentzia automatikoki esleitzeko. Sentimenduen lexikoiak hitzei esleitzen die sentimendu-balentzia, eta ondoren, sentimenduen sailkatzaileak eragiketa egiten du esaldiaren sentimendu-balentzia kalkulatzeko. Hurrengo adibideek hori erakusten dute.

- (41) Halere, pentsa litekeenaren aurka, gaien urritasunak eta diskurtso [errepikakorrak]₋₆ ez dakarte ñabardura aberastasunik, are gutxiago argumentu-mailako sakontasunik.[₋₆] (LIB18)
- (42) Arazo [nagusia]₊₂, nire ustez, gaien [emankortasun]₊₄ zalantzazkoan eta ekintzaren bilakaera [eskasean]₋₃ datza.[₊₃] (LIB18)
- (43) (...) Emaidza [ezustekorik]_{-1.5} gabeko istorio bat da, irakurlea [epel]_{-1.5} uzteko arrisku dezente duen tonu arras moderatu batean emana.[₋₃] (LIB18)

Goiko adibideetan, hitzek eskuinetara sentimendu-balentzia bat dute, euskarazko sentimenduen sailkatzailean inplementatu dugun sentimenduen lexikoiarenak. Esaldien amaieran, esaldi osoaren sentimendu-balentzia ageri da eta esaldian dauden sentimendu-balentzi guztien batura da. Lehen adibidean, esaterako, lexikoiaren arabera hitz batek, *errepikakorrak* hitzak, bakarrik du sentimendu-balentzia -6 eta ondorioz, esaldiarena ere halaxe da⁵. Hurrengo bi adibideetan ere, lexikoiak funtzionamendu bera du euskarazko sentimenduen sailkatzailearen baitan.

4.2.2. Lexikoiaren sorkuntza

Euskarazko sentimenduen lexikoia sortzeko metodologian, hartu ditugun erabakiak azalduko ditugu (4.2.2.1 atala); ondoren, gaztelaniazko lexikoia euskaratzerakoan agertu zaigun kasuistikaren berri emango dugu (4.2.2.2 atala);

⁵Gogoan hartu behar da lexikoiko sarreraren sentimendu-balentzia -5 eta $+5$ artekoa dela, baina SO-CAL tresnan badaude zenbait eragiketa hizkuntzarekin zerikusia dutenak eta hizkuntza guztietan aplikatzen direnak eta horiek hitzaren sentimendu-balentzia indar edo ahul dezakete. Kasu horretan, *errepikakorrak* hitzaren sentimendu-balentzia -6 raino indartu da.

jarraian, gaztelaniazko lexikoi horren euskarazko ordainak ematen sentimenduen balentziak aukeratzeko zein irizpide hartu ditugun kontuan azalduko dugu (4.2.2.3 atala).

4.2.2.1. Metodologian hartutako erabakiak

Sentimenduen lexikoiko hitzek eta beren sentimendu-balentziek elkarren artean koherentzia izan dezaten, metodologian zehar erabaki hauek hartu ditugu:

- Nahiz eta gaztelaniazko lexikoa euskaratu dugun, kasu gehienetan, euskarazko ordainei ingelesezko lexikoia sentimendu-balentziak esleitu dizkiegu. Lan honetan gaztelania edota ingelesa izan zitezkeen hastapeneko hizkuntzak. Guk gaztelania hobetsi dugu hastapeneko hizkuntza moduan arrazoi hauengatik: alde batetik, sentimenduen lexikoia gaztelaniako bertsioaren batez besteko doitasuna % 71,81ekoa da, ingelesekoarena % 76,65 den bitartean. Beste aldetik, gaztelaniako hitzei euskarazko ordaina emateko baliabide gehiago daude ingelesetik baino eta, gaztelaniatik ordaina ematea errazagoa izango litzateke.

Halaber, beste arrazoi bat badago eta hori hizkuntzari lotuta dauden ezaugarri soziokulturalak dira. Euskarak eta gaztelaniak ezaugarri soziokultural gehiago partekatzen dituzte, ingelesak eta euskarak baino. Horren adibide dugu 4.2.2.2 azpiataleko 4. Fenomenoa, *frankismo* hitzaren kasua, hain zuzen ere. Euskarak eta gaztelaniak partekatzen duten ale hitz bat da. Horrelako kasu gehiago egon dira euskarazko ordainak ematean. Hizkuntzen artean gertatzen diren itzulpeneko kasu hauek Meng *et al.*en (2012) lanean aipatzen dira eta Mohammad *et al.*ek (2016) itzulpenetan gertatzen diren sentimenduen aldaketak ere aipatzen ditu.

Bestalde, ingelesezko lexikoia gaztelaniazko lexikoa baliatuta euskaraz sortu dugun lexikoian esleitu diren balentziak egokiak diren aztertze eta ez badira zuzentzeko balio izan digu. 3.10 taulako 3. Fenomenoan (*seinale*) gertatu den moduan, euskarazko ordain batek aukeran gaztelaniazko lexikoiko ale hitzaren (*cicatriz*) eta ingelesezko

lexikoiko ale hitzaren (*signal*) sentimendu-balentziak dituenen, ingeleseko sentimendu-balentziaren alde egin dugu, batez besteko doitasuna handiagoa duelako (% 76,65). Euskarazko sarrerak ingeleseko lexikoian ordainik ez duenean, aldiz, kasu batzuetan euskarazko sarrera hori lexikoitik kendu egin dugu, gaztelaniako lexikoaren doitasun baxuagoak (% 71,81) euskarazko lexikoaren kalitatea jaitsi dezakeelakoan. Ondorioz, 576 gaztelanizko hitzei, hasiera batean, euskarazko ordaina eta sentimendu-balentzia esleitu diegun arren, horiek euskarazko sentimenduen lexikoitik kendu egin ditugu, ingelesezko lexikoian agertzen ez direlako.

Beraz, batetik, lexikoaren gaztelaniazko bertsioaren estaldura eta ezauzgarri soziokulturalak eta, bestetik, lexikoaren ingelesezko bertsioaren doitasuna (hitzen balentzia) uztartu ditugu.

- Gaztelaniazko lexikoiko hitz baten euskarazko hainbat ordain kontuan hartzea. Gaztelaniazko lexikoiko hitzei euskarazko ordaina ematean izan dugun beste zalantzetako bat euskarazko ordain guztiak kontuan hartu behar diren izan da. Izan ere, gaztelaniazko hitzak batzuetan euskarazko hainbat ordain izan ditzke esanahi bera mantenduz. Esaterako, gaztelaniazko hitz batek euskarazko hainbat izan ditzake euskararan, 4.7 taulan ikusten den moduan. Halaber, hitzak polisemikoak ere izan daitezke, eta horrek ere gaztelaniazko hitz bati euskarazko hainbat ordain ematea eragiten du; nahiz eta esanahia kasu horretan ezberdina den. Egoera horren aurrean, hau da, gaztelaniazko hitz bati euskarazko hainbat ordain ahal zaizkionean, euskarazko ordain posible guztiak aintzat hartzea erabaki dugu sortuko den lexikoia ahalik eta estaldura handiena izan dezan.

4.7 taulan bi adibide daude non ordain posible guztiak kontuan hartu direla erakusten den. Gaztelaniazko lexikoian *enfermedad* eta *horror* hitzak daude eta horiek *Elhuyar* (Elhuyar, 2013) eta *Zehazki* (Sarasola, 2005) hiztegiak erabiliz euskarazko ordainak lortu dira eta hitz horietako bakoitzak euskararan hainbat ordain posible ditu, hirugarren zutabearen agertzen den moduan.

- Hitz polisemikoen domeinuaren egokitzapena. Gaztelaniako lexikoiko

Gaztelaniaz jatorrian	Balentzia	Ordain posible guztiak euskaraz
enfermedad (izena)	-1	gaixotasun, gaitz, eritasun, afekzio, ezontsa
horror (izena)	-4	laztura, izu, lazgarrikeria, izugarrikeria

4.7 taula: Euskarazko ordain posible bat baino gehiago dituzten adibideak.

hitzei euskarazko ordaina ematean eta sentimendu-balentzia esleitzean, testuinguruaren arabera esanahi ezberdinak eta zeinu ezberdinetako (+ edo -) sentimendu-balentziak dituzten ordainak agertu dira, 4.8 taulako modukoak. Egokiena euskarazko ordain guztiak (eta sentimendu-balentziak) kontuan hartzea izango litzateke, nahiz eta euskarazko ordainek esanahi ezberdinak izan, baina horrek lexikoia konplexutasuna asko handituko luke. Horretaz gain, lexikoa inplementatuta egongo litzakeen sistemak hitzen adiera-desanbiguaziorako gaitasuna eduki beharko luke.

Hori konpontzeko, esanahi ezberdinak eta zeinu ezberdinetako sentimendu-balentziak dituzten ordainetan esanahi bat bakarrik hartzea erabaki dugu. Hautaketa egitean, Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) agertzen diren testuak eta domeinuak erabili ditugu.

4.8 taulan, hitz polisemikoetan balentzia hautaketa nola egin den esplikatzen da. *Deigarri* eta *lehiatsu* adjektiboek esanahi ezberdinak (*deigarriren* kasuan, *aparatoso* -3, *llamativo* +3 eta *lehiatsuren* kasuan, *ansioso* -3, *apasionado* +4 eta *competitivo* +3) dituzten hitzen ordainak dira eta, metodologiako laugarren urratsean, euskarazko ordain horiei sentimendu-balentzia esleitu behar diegu. Aukeran sentimendu-balentzia bat baino gehiago daudenez (*deigarriren* kasuan, -3 eta +3 eta *lehiatsuren* kasuan, -3, +4 eta +3), euskarazko ordain horien testuingurua aztertu dugu Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) gako-hitzak testuinguruan (*Key Word In Context*, KWIC) erabiliz eta ordainen corpuseko testuinguruekin hobekien datorren sentimendu-balentzia aukeratu dugu. *Deigarri* eta *lehiatsu* ordainetan, +3 sentimendu-balentzia izan da hautatu duguna.

- Koherentziaren sendotasuna. Euskarazko ordaina lortu ondoren, ordai-

Euskal ordaina	Jatorriz gaztelaniaz eta balentzia	Bal. aukeratua
deigarri (adj.)	aparatoso -3 , llamativo +3	+3
lehiatsu (adj.)	ansioso -3 , apasionado +4 , competitivo +3	+3

4.8 taula: Hitz polisemikoen tratamendua.

nei sentimendu-balentzia esleitu diegu, balentziak bat zetozen kasuetan, balentzietan koherentzia mantentzen saiatu gara.

Irizpidea	Euskara	Balentzia	Euskara	Balentzia
A	errukigabe	-4	errukigabeko	-4
B	tonto	-3	tuntun	-3
C	arduradun	+2	arduragabe	-2

4.9 taula: Itzulpen-prozesuaren koherentzia erakusten duten adibideak.

4.9 taulan ageri dira zenbait adibide. Bertan agertzen diren kasuak euskarazko ordaina ematerakoan laugarren urratsekoak dira, balentzia hautatu behar den mementokoak. Guztira hiru kasu ezberdin antze-man ditugu:

- A- Izenlagunarekin agertzen diren euskarazko ordainak. 4.9 taulan, A kasuan, *errukigabe* lema ageri da. *Errukigabe* eta *errukigabeko* izenondoak dira, baina, bigarrenak izenlagunaren *-ko* atzizkia ere badu. Bi hitzei balentzia hautatzeko mementoan, balentzia bera eman diegu; gramatika-kategoria gorabehera, esanahi bera baitute. Aipagarria da horrelako kasu asko egon direla balentzia hautaketan zehar. Izenlaguna daraman euskarazko ordainaren beste kasu bat *berehala/berehalako* hitz pareta izan da eta biei +2 balentzia esleitu diegu. Kontuan hartu behar da, kasu hori euskarazko lexikoia lehene bertsioa sortzerakoan gertatu dela, lexikoia bigarren bertsioan hiztegi-tako sarrerak diren hitzak bakarrik hartu baititugu kontuan.
- B- Esanahi bera, baina genero ezberdina. Kasu gutxi batzuk egon dira non bi hitz ezberdin erabiltzen diren generoarengatik, bai-

na esanahi bera dutenak. Horrelako kasuetan ere bi hitzetan sentimendu-balentzia bera jartzen ahalegindu gara; aukeran zeuden sentimendu-balentziek uzten zuten heinean. Hala, 4.9 taulako B kasuan, gizonaezkoentzat *tonto* eta emakumezkoentzat *tuntun* hitzak ditugu, aipaturiko egoera islatzen dutenak. Ezberdintasun bakarra hitzek erreferentzia egiten dioten pertsonen generoa denez, biei sentimendu-balentzia bera (-3) esleitu diegu.

C- Hizkiak dituzten hitzak. Gaztelaniako lexikoiko hitzei euskarazko ordaina ematean agertu den beste hitz-multzo mota da hizkiak dituztenena. Hizkiak erabiltzearengatik, aurkakotasuna adierazten duten hitzen bikoteak agertu dira ordaina ematean. Hitz bat aurkakotasuna adierazten duen hizkiarekin eta hori gabe agertzea nahiko ohikoa izan da. Aurkakotasuna adierazteko gehien agertu diren hizkiak ezeztapenarekin lotutako hauek izan dira: *des-*, *ez-*, *-ezin* eta *-gabe*. Hitz bera aurkakotasun hizkiarekin eta gabe agertu izan denean, bi egoeretako intentsitate bereko sentimendu-balentzia bera jarri diegu, baina zeinu ezberdinarekin. 4.9 taulan, C kasuan, hitz bera ageri da (*ardura*), baina batean *-dun* atzizkia du, jabetza adierazten duena eta bestean, berriz, *-gabe* atzizkia du, gabezia adierazten duena.

- Ordain *ez-zuzenak*. Euskarazko ordainak ematean agertu den beste fenomenoetako bat ordain *ez-zuzenena* da. Ordain horiek berez zuzenak dira, baina gure ikerketa-lanerako ez dira baliagarriak testuinguru zehatz batzuetan erabiltzen direlako edota Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) domeinuekin ez dutelako loturarik.

Adibidez, gaztelaniazko lexikoian agertzen den *provinciano* (-1) hitza, *atzeratu* moduan itzulita zuzena eta erabilgarria izango litzateke guretzat; izan ere, mespretxuzko esanahi bat da eta horiek erabilgarriak dira sentimenduen analisisian (Sorby, 2008). Hala ere, *Elhuyar* hiztegia (Elhuyar, 2013) erabilita *Araba*, *Bizkaia* edo *Gipuzkoako herritar* (*gipuzkoarra bereziki*) izan da lortu den ordainetako bat. Ordain hori gure lanerako ez da erabilgarria. Horrenbestez, ordain hori ez dugu aintzakotzat hartu.

- Hizkuntza figuratiboa. Aurkitu dugun beste ordain-mota bat hizkuntza figuratiboa da. Bertan, hitzek edo hitz-multzoek ez dute berez duten esanahia. Hitzunak esanahia moldatu egiten du, bigarren mailako esanahia emanez (Ghosh *et al.*, 2015). Bigarren mailako esanahi horiek zuzenak dira, baina testuinguru jakin eta mugatu batean erabiltzen direnez, horrelako ordainak kontuan ez hartzea erabaki dugu.

Adibidez, gaztelaniazko lexikoiko hitzari euskarazko ordaina ematean *beltza* (-2) azaldu da. Hitz horrek bi esanahi ditu: i) beltza, kolorea eta ii) tristea, zoritxarrezkoa. Azken hori esanahi figuratiboa da, bigarren mailako esanahia. Hitzari sentimendu-balentzia aukeratzeko urratsean, euskarazko ordainak gaztelaniazko hitzaren -2 balentzia du jaso du; gaztelaniazko hitzak esanahi figuratiboari egiten diolako erreferentzia. Nahiz eta, hori egokia izan arren, gure lexikoian ez sartzea erabaki dugu: alde batetik, hitzaren erabilera hori ez delako ohiko esanahiarena baino handiagoa eta, beste aldetik, gure corpusera begira ere ez da egokia.

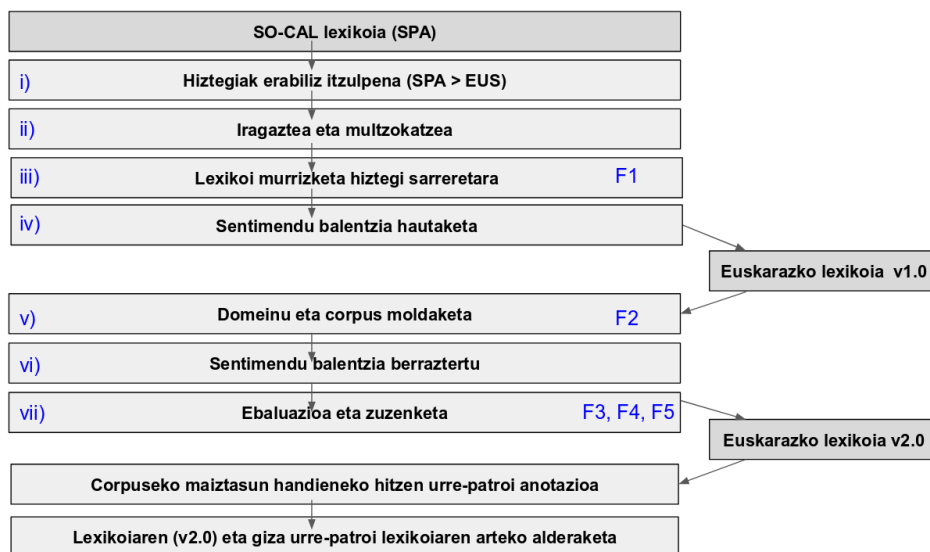
4.2.2.2. **Gaztelaniazko lexikoiko hitzari euskarazko ordaina ematean egondako kasuistikak**

SO-CAL tresnaren (Taboada *et al.*, 2011) gaztelaniazko lexikoa euskaratzerakoan, 3.1.2 ataleko metodologiari jarraitu dugu. Hala ere, gaztelaniazko hitzei euskarazko ordaina ematean, gaztelaniazko hitz guztiek ez dute amaiera bera izan. Hau da, euskarazko ordaina eman beharreko hitzaren ezaugarrien arabera, urrats gehiago edo gutxiago bete behar izan dira gaztelaniazko hitzei ordaina emateko.

4.3 irudian azaltzen dira SO-CALen gaztelaniako bertsioko lexikoiko hitzei euskarazko ordainak emateko urratsak⁶ eta euskarazko ordaina ematean agertu diren bost fenomenoak. “Fenomeno”⁷ hitza aipatzean, gaztelaniazko lexikoiko hitzek urratsei dagokienean hasiera bera, baina amaiera ezberdina izan dutela adierazi nahi dugu. F1-F5 bidez dago adierazita euskarazko ordainetako bakoitzak non bukatzen duen bere ibilbidea.

⁶4.3 irudian, urratsak zenbaki erromatarren bidez adierazita daude.

⁷4.3 irudian, fenomenoak Fren bidez adierazita daude.



4.3 irudia: Gaztelaniako lexikoiko hitzei euskarazko ordaina emateko gauzatutako urratsak.

Jarraian, bost fenomeno horietako bakoitza esplikatuko dugu egitura honi jarraituz: lehenik eta behin, fenomeno zertan den azalduko dugu; ondoren, fenomeno horren atzean dagoen arrazoia aditzera emango dugu eta, azkenik, fenomeno horren adibide bat emango dugu.

- Hiztegi-tako sarrerak ez diren euskarazko ordainak (F1).
 - Fenomenoa. SO-CAL tresnaren gaztelaniako lexikoiko (Brooke *et al.*, 2009) hitzari euskarazko ordaina eman diogu da, baina ordaina ez denez *Elhuyar* (Elhuyar, 2013) eta *Zehazki* (Sarasola, 2005) hiztegi-tako sarrera bat ez dugu aintzakotzat hartu.
 - Arrazoia. Sortzen ari garen lexikoiko sarrerek ezaugarri berak izatea nahi dugu eta, horregatik, lexikoian zer sartu erabakitzerakoan muga bat jarri behar izan dugu eta muga hori euskarazko ordaina hiztegi-tako sarrera bat izatea da.
 - Adibidea. Gaztelaniako *desacreditar* hitzari hiru euskarazko ordain (*ospea_kendu*, *izena_kendu* eta *sona_kendu*) eman dizkiogu eta

ordainetako bakoitzak jatorrizko hitzaren balentzia -2 oinordekotzan hartu du. Baina, ordain horiek ez dira hiztegi-tako sarrerak eta, horrenbestez, lexikoitik kendu egin ditugu.

- Corpusean agertzen ez diren euskarazko ordainak (F2).
 - Fenomenoa. SO-CAL tresnaren gaztelaniako lexikoiko (Brooke *et al.*, 2009) hitzari euskarazko ordaina eman diogu, euskarazko ordaina *Elhuyar* (Elhuyar, 2013) edota *Zehazki* (Sarasola, 2005) hiztegi-tako sarrera bat da, baina ordaina ez da Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) agertzen. Horrenbestez, hitz hori ez dugu kontuan hartu euskarazko lexikoia-aren bigarren bertsioan⁸.
 - Arrazoa. Corpuseko domeinuekin loturarik ez duten hitzak lexikoitik kendu egin ditugu. Modu horretan, lexikoia-ko corpusarekiko koherentzia izatea nahi dugu. Erabaki horren ondorioz, lexikoia- lehen bertsioan 8.140 sarrera izatetik bigarren bertsioan 1.813 sarrera izatera igaro da.
 - Adibidea. Esaterako, *atrofiatu* hitzak -1 balentzia du, baina hitz hori ez da Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) azaltzen. Izan ere, aztertzen ari garen hitza ez da corpusean dauden domeinuetan ohikoa. Hitza corpusean agertzen ez denez, lexikoitik kendu egin dugu; (corpuseko) domenu-*ei* loturiko lexikoia-eratu nahi baitugu.
- Balentziaren egokitasuna zalantzarikoa duten euskarazko ordainak (F3).
 - Fenomenoa. SO-CAL tresnaren gaztelaniako lexikoiko (Brooke *et al.*, 2009) hitzei euskarazko ordaina eman diegu, ordaina *Elhuyar* (Elhuyar, 2013) edota *Zehazki* (Sarasola, 2005) hiztegi-tako sarrera bat da eta ordaina Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016)

⁸4.3 irudian, euskarazko lexikoia-aren lehen bertsioa iv) eta v) urratsen artean ageri da eta bigarren bertsioa vii) urratsaren ondoren. Euskarazko lexikoia-aren lehen bertsioan, gaztelaniako lexikoiko hitz guztiei eman diegu euskarazko ordaina, baina euskarazko lexikoia-aren bigarren bertsioan, Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) agertzen ez diren euskarazko ordainak kendu ditugu.

agertzen da. Baina, ordainak duen balentziaren egokitasuna zalantzazkoa da eta zalantza argitzeko SO-CAL tresnaren (Taboada *et al.*, 2011) ingeleseko bertsioaren lexikoian agertzen den egiazta-tu da eta ez da azaltzen. Ondorioz, lexikoia-aren bigarren bertsioan, hitz hori kendu egin dugu, baina lehen bertsioan mantentzen du-gu.

- Arrazoa. Euskarazko sarreraren baliokiderik ez dago SO-CALen (Taboada *et al.*, 2011) ingelesezko lexikoian eta gaztelaniazko ber-tsiotik lortutako balentzia ez da egokitzat jotzen, Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) dauden domeinuak kontuan har-tuta. Hori dela eta, euskarazko sarrera lexikoitik kentzea erabaki dugu.
- Adibidea. Fenomeno horren adibidea da *seinale* hitza. SO-CALen gaztelaniako lexikoitik (Brooke *et al.*, 2009) -1 balioa hartu du zaio, baina ez dugu egokitzat jo, nahiz eta balentzia intentsitate txikikoa izan. Arrazoi horregatik, hitza lexikoitik kentzea erabaki dugu.
- Arrazoi soziokulturalengatik aintzat hartutako euskarazko ordainak (F4).
 - Fenomenoa. SO-CAL tresnaren gaztelaniako lexikoiko (Brooke *et al.*, 2009) hitzari euskarazko ordaina eman diogu, ordaina *Elhuyar* (Elhuyar, 2013) edota *Zehazki* (Sarasola, 2005) hiztegi-tako sar-rerra bat da, ordaina Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) agertzen da, baina ordain hori ez da SO-CALen (Taboada *et al.*, 2011) ingeleseko lexikoiko hitz bat. Kalitatea zalantzaz-koa ez denez, nahiz eta ingeleseko lexikoian ez agertu, lexikoia-aren lehen bertsioan nahiz bigarre-nean jaso dugu.
 - Arrazoa. Euskarazko sarreraren baliokiderik ez dago SO-CALen (Taboada *et al.*, 2011) ingeleseko lexikoian eta gaztelaniazko ber-tsiotik lortutako balentzia egokitzat jotzen da, Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) dauden domeinuak aintzakotzat hartuz. Hau da, hitz hori ingelesean ez da hain ohikoa arrazoi soziokulturalengatik, baina bai gaztelanian. Horregatik eta kali-

tatea zalantzazkoa ez denez, euskarazko ordaina aintzakotzat hartzen da. Ondorioz, hitz bera aurreko fenomenoaren egoera berean egon arren, emaitza ezberdina da eta hitza lexikoian mantentzea erabaki dugu.

- Adibidea. Esaterako, euskarazko lexikoiko sarrera bat *frankismo* da eta -2 balioa hartu du gaztelaniazko bertsioetik. Hitz hori espainiar politikari lotutakoa da eta, horregatik, gaztelaniazko bertsioan agertzen da, baina ingelesezkoan ez. Hitzaren ezaugarri hori kontuan hartu dugu eta, horregatik, euskarazko lexikoian mantentzea erabaki dugu.
- Gaztelaniazko eta ingelesezko lexikoietatik eta corpusetik elikatutako euskarazko ordainak (F5).
 - Fenomenoa. SO-CAL tresnaren gaztelaniako lexikoiko (Brooke *et al.*, 2009) hitzari euskarazko ordaina eman diogu, ordaina *Elhuyar* (Elhuyar, 2013) edota *Zehazki* (Sarasola, 2005) hiztegieta sarrera bat da, ordaina Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) agertzen da, baita SO-CALen (Taboada *et al.*, 2011) ingelesezko lexikoian ere. Egoera horretan, hitza euskarazko lexikoian bi bertsioetan dago.
 - Arrazoia. Euskarazko sarreraren baliokidea badago ingelesezko bertsioan ez ezik, gaztelaniazkoan ere badago, eta euskarazkoari balentzia esleitzerakoan ingelesezko bertsioa hobesten da.
 - Adibidea. 3.10 taulako kasuan, *zuzen* hitzak gaztelaniazkoetik +3 balentzia oinordekotzan hartu du, eta ingelesezkoan bi aukera daude: *right* +1 balentziarekin eta *correct* +3 balentziarekin. Gaztelaniazko hitzak (*correcto*) eta ingelesezko hitzak (*correct*) balentzia bera dutenez (+3), *zuzen* hitzak gaztelaniatik oinordekotzan hartutako balentzia mantentzea jo dugu egokituz.

4.2.2.3. Euskarazko ordainari balentzia aukeratzeko irizpideak

Euskarazko ordainari dagokion sentimendu-balentzia eta gaztelaniako esanahia aukeratzekoan, jarraian azaltzen den prozedura jarraitu dugu.

- Prozedura 1.
 - Azalpena. Baldin eta jatorrizko gaztelaniazko hitzak euskarazko ordain bakarra eta egokia badu, euskarazko ordainak haren sentimendu-balentzia hartuko du.
 - Adibidea. Hori gertatu da 2. eta 4. Fenomenoetan. Fenomeno horietan, ordain bakarra dago eta ordain bakar hori egokia da, ez delako ez esanahi figuratiboa, ez Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) domeinuekin bat ez datorren ordaina. Hori dela eta, *atrofiatu* hitzari -1 balentzia esleitu diogu eta *frankismo* hitzari -2 balentzia.

- Prozedura 2.
 - Azalpena. Baldin eta euskarazko ordainak jatorrian hainbat gaztelaniazko hitz eta haien balentziak baditu, hitzari esleituko diogun sentimendu-balentzia (eta gaztelaniazko esanahia) Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) oinarrituta egingo dugu. Lan hori corpusean oinarrituta dagoenez, gaztelaniazko esanahien (eta sentimendu-balentzien) artetik aukeraketa egin beharko dugu eta aukeratutakoak bat etorri edo egokia izan beharko Euskarazko Iritzi Corpusean (Alkorta *et al.*, 2016) hitz hori agertzen den testuinguruarekin. Hori gauzatu ahal izateko hitz-gakoa testuinguruan (*Key Word in Context*, KWIC) erabili dugu.
 - Adibidea. Prozedura hori jarraitu dugu *erakargarri* hitzaren kasuan. Lexikoiko gaztelaniako hitzei euskarazko ordaina eman ondoren, *erakargarri* hitzaren jatorrian hitz eta sentimendu-balentziak hauek daude: *apasionante* +5, *apetecible* +2, *apetitoso* +2, *atractivo* +2, *fascinante* +5, *glamuroso* +4, *goloso* +3, *interesante* +4, *irresistible* +4, *seductor* +1 eta *tentador* +2. KWIC teknikaren bidez, *erakargarri* hitzaren testuingurua corpusean behatu ondoren, +2 balentzia esleitu diogu, corpuseko testuinguruarekin bat gehien berak egiten duelako.

- Prozedura 3.

- Azalpena. Euskarazko lexikoaren lehen bertsioa euskaraz sortzerakoan, kasuak egon dira non gaztelaniazko hainbat hitzek euskaraz ordain bera duten, eta euskarazko ordain hori corpusean ez den agertzen. Egoera horretan, euskarazko ordainak esanahirik zabalena duen gaztelaniazko hitzaren sentimendu-balentzia esleitu diogu.
- Adibidea. *Bat_uz_datorrena* hitzaren kasua prozedura horren adibidea da. Egitura hori ez da corpusean agertzen; eta, beraz, bere aukeretan (*descontento* -3, *discordante* -2, *insatisfecho* -2) esanahirik erabilienaren (*discordante*) balentzia (-2) hautatu dugu.

4.2.3. Euskarazko sentimenduen lexikoiaren ebaluazioa

Hasteko, 4.2.3.1 azpiatalean, ebaluazioa egiteko prozedura aurkeztuko dugu. Ondoren, 4.2.3.2 azpiatalean, ebaluazio horren emaitzak azalduko ditugu.

4.2.3.1. Ebaluazioa egiteko prozedura

Euskarazko sentimenduen lexikoa ebaluatzeko honako baliabideak erabili ditugu:

- 400 hitzeko zerrenda. *Analhitza* (Otegi *et al.*, 2017) tresna baliatuz, Euskarazko Iritzi Corpusetik (Alkorta *et al.*, 2016) gramatika-kategoria bakoitzeko (izenak, adjektiboak, adberbioak eta aditzak) maiztasun handieneko 100 hitz lortu ditugu.
- Urre-patroia. Horretarako, bi etiketatzailek 400 hitzei banan-banan -5 eta +5 arteko sentimendu-balentzia esleitu die.

Euskarazko sentimendu lexikoiak zerrendako 400 hitzi sentimendu-balentzia esleitu die automatikoki eta, ondoren, sentimendu-balentzia horiek urre-patroiko sentimendu-balentziekin alderatu dira. Bi horien arteko adostasuna neurtzeko Pearson korrelazioa (Benesty *et al.*, 2009) erabiltzea erabaki dugu. Etiketatzaile edo lexikoi batek hitz bati -3 balentzia esleitzen dion bitartean, beste etiketatzaile edo lexikoi -1 edo +1 balentzia esleitzen badio, Pearson korrelazioan oinarrituta aintzat hartuko dugu -1 balentziak +1 balentzia baino

hurbiltasun handiagoa duela eta, horrenbestez, bi etiketatzeren arteko adostasuna handiagoa dela. Gainera, kalitatearen neurketa bi modutan egitea erabaki dugu jarraian azaltzen diren bi arrazoiengatik:

- Pearson 1: bi etiketazaileek etiketatutako hitzak bakarrik kontuan hartuta egindako korrelazio neurketa da. Beraz, 400 hitz horietatik ez dira guztiak erabili neurketa egiteko mementoan.

4.10 taulan, Pearson 1 deitu dugun korrelazioaren neurketa nola egin den ikus daiteke. Taulan 10 hitz daude, baina horietako hiru (*ageri*, *ar* eta *bakoitza*) ez daude bi etiketazaileek etiketatuta. Horrenbestez, Pearson korrelazioaren koefizientea kalkulatzean, hiru hitz horiek ez dira kontuan hartu.

Adjektiboak	Eti1	Etik2	Etik1	Etik2
ageri	POS		1	
ahul	NEG	NEG	3	3
antisozial	NEG	NEG	1	5
apur	NEG	NEG	1	1
ar				
argi	POS	POS	2	3
aspergarri	POS	NEG	3	3
ausart	POS	POS	3	4
bakar	NEG	POS	1	5
bakoitz				

4.10 taula: Bi etiketatzaileraren arteko korrelazioaren kalkulua, etiketatatu gabeko hitzak aintzat hartu gabe (Pearson 1).

- Pearson 2: corpusetik ateratako hitz guztiak, 400 hitz guztira, erabili dira. Batzuetan, bi anotatzaileek hitzei sentimendu-balentzia jarri diete. Beste kasu batzuetan, aldiz, anotatzaileak edo anotatzaileek ez dio(te) hitzari sentimendu-balentzia jarri, haien ustez ez duelako sentimendu-balentziarik behar. Horrelako kasuetan, anotatu gabeko hitzei 0 sentimendu-balentzia esleitu diegu. Hori da Pearson 1ekiko ezberdintasun bakarra, anotatu gabeko hitzei 0 sentimendu-balentzia esleitzea. Modu horretan, hitz horiek ere neurketan erabili ditugu.

4.11 taulan, Pearson 2 deitu dugun korrelazioaren neurketa nola egin den azaltzen da. 4.10 taulan bezala, anotatzaileek hiru hitzi (*ageri*, *ar* eta *bakoitza*) ez diete sentimendu-balentzia esleitu. Baina horiek ere neurketan kontuan hartu nahi ditugu eta, horregatik, 0 sentimendu-balentzia esleitu diegu.

Adjektiboak	Etik1	Etik2	Etik1	Etik2
ageri	POS	0	1	0
ahul	NEG	NEG	3	3
antisozial	NEG	NEG	1	5
apur	NEG	NEG	1	1
ar	0	0	0	0
argi	POS	POS	2	3
aspergarri	POS	NEG	3	3
ausart	POS	POS	3	4
bakar	NEG	POS	1	5
bakoitz	0	0	0	0

4.11 taula: Bi etiketatzaileraren arteko korrelazioaren kalkulua, etiketatatu gabeko hitzak ere aintzat hartuz (Pearson 2).

4.2.3.2. Ebaluazioaren emaitza

Jarraian, egin ditugun bi ebaluazioak aurkeztuko dugu. Lehenik eta behin, bi etiketatzaileraren arteko korrelazioa azalduko dugu. Ondoren, urre-patrioiaren eta euskarazko sentimenduen lexikoiaren arteko korrelazioaz arituko gara. Zerrendako hitzei sentimendu-balentzia esleitzerakoan dagoen bi pertsonen arteko korrelazioa hurrengo 4.12 taulan ikus daiteke.

Gramatika-kategoria	Pearson 1	Pearson 2
Izena	0,87	0,59
Adjektiboak	0,71	0,60
Aditzondoak	0,93	0,82
Aditzak	0,87	0,76
Guztira	0,79	0,73

4.12 taula: Bi anotatzaileen arteko Pearson korrelazioaren neurketa.

Pearson 1 balioak erakusten korrelazio-koefizientea altua dela (0,79). Horrek bi anotatzaileek zerrendako hitz askotan hitzei esleitu dieten sentimendu-balentzia nahiko antzekoa dela esan nahi du. Gramatika-kategorien artean badaude ezberdintasun batzuk, korrelaziorik altuena 0,93koa delako eta baxuena, aldiz, 0,71koa. Korrelaziorik altuena aditzondoei dagokie eta baxuena adjektiboei. Pearson 2ri dagokionez, emaitzek erakusten dutenez, korrelazio-koefizienteak altua izaten jarraitzen du, nahiz eta Pearson 1ekin alderatuta zerbait baxuagoak diren. Pearson 2ren kasuan, gramatika-kategorietan koefizienteak 0,82 eta 0,60 artean kokatzen dira.

Kontingentzia-taulak emaitzak interpretatzeko informazio osagarria ematen digu 4.13 taulan. Kontingentzia-taula hori Pearson 2rena da; izan ere, zutabe eta zerrendetan 0 agertzen da, eta horrek adierazten du anotatzaileek hitzari ez diotela sentimendu-balentzia esleitu. Bertan ikusten denez, bi anotatzaileen arteko ezberdintasuna nagusiki egoera batean gertatzen da: anotatzaile batek hitzari balentzia esleitzen dionean besteak ez dio esleitzen, eta alderantziz. Egoera hori desadostasun guztien % 90,19ren jatorrian dago (ezberdintasuna dagoen 102 instantzietatik 92etan, hain zuzen ere).

Guztira kategoriak			
	0	Negatiboa	Positiboa
0	187	12	27
Negatiboa	14	42	5
Positiboa	39	5	69

4.13 taula: Bi anotatzaileen arteko kontingentzia-taula.

4.14 taulan, berriz, sentimenduen lexikoiaren eta urre-patroiaren arteko korrelazio-koefizienteak ageri dira. Urre-patroiaren eta sentimenduen lexikoiaren arteko korrelazio-koefizienteen neurketak ezberdintasun batzuk erakusten ditu aurretik egindako bi anotatzaileen arteko korrelazio koefizientearekiko.

Pearson 1i dagokionez, hau da, sentimendu-balentzia esleitu zaien hitzak bakarrik kontuan hartuta, korrelazio-koefizientea aurreko neurketatik gertu dago. Kasu horretan, koefizientea 0,76 da eta aurreko neurketan 0,79. Koefizientean ezberdintasun handiak daude gramatika-kategorien artean. Koefizienterik altuena izenetan dago (0,96) eta baxuena, berriz, aditzetan (0,69).

Gramatika-kategoria	Pearson 1	Pearson 2
Izena	0,96	0,59
Adjektiboak	0,78	0,56
Aditzondoak	0,75	0,47
Aditzak	0,69	0,54
Guztira	0,76	0,54

4.14 taula: Euskarazko sentimenduen lexikoa (V2.0) eta urre-patroiaren arteko Pearson korrelazioaren neurketa.

Baina, ezberdintasun esangurantsua Pearson 2n gertatzen da, bi anotatzaielen arteko korrelazio koefizientearekin alderatzen denean. Pearson 2an koefizientea 0,54 izan da; aurreko neurketan 0,73 izan denean (0,19ko ezberdintasuna). Koefizienterik altuena izenenak izaten jarraitzen du (0,59) eta baxuena, berriz, aditzondoena da (0,47).

Datu hauen gure interpretazioa hauxe da: lexikoiaren eta urre-patroiaren arteko Pearson 1 koefizienteak altua izaten jarraitzen du zerrendako hitzei esleitu zaien balentzia antzekoa izan delako bi kasuetan, batean, bi anotatzaielen artean eta bestean, urre-patroia eta lexikoiaren artean. Baina, ezberdintasunak daude balentzia esleitu zaien zerrendako hitzetan. Bi anotatzaielen kasuan, balentzia esleitu zaien hitzak kasik berak izan dira. Hemen, urre-patroiaren eta lexikoiaren arteko emaitzak kontuan hartuz, balentzia esleitu zaien hitzak beti ez dira berak izan.

Kontingentzia-taulak lexikoiaren eta urre-patroiaren arteko ezberdintasunak argitara ematen ditu 4.15 taulan. Bi etiketatzaileen artean ezberdintasunak balentzia esleitu beharreko hitzaren inguruan daude. Batek hitz bati balentzia esleitzen dionean, besteak ez egitea, eta alderantzizko kasua. Lexikoiak edo urre-patroiak hitz bati balentzia esleitzean besteak balentzia ez esleitzea ezberdintasun guztien % 89,83 (118tik 106 kasutan) izan da. Bi anotatzaielen hitzei sentimendu-balentzia esleitzean ere gertatu da anotatzaielen batek hitzari balentzia esleitzean beste anotatzaielen ez esleitzea, baina askoz intentsitate baxuago batean, kasu horretan ez bezala. Baina, bi pertsonen arteko emaitzak alderatuta badago beste ezberdintasun bat. Bi pertsonen artean, batek balentzia esleitzean besteak ez egitea, eta alderantziz, nahiko

Guztira kategoriak			
	0	Negatiboa	Positiboa
0	195	2	15
Negatiboa	30	34	8
Positiboa	59	4	53

4.15 taula: Euskarazko sentimenduen lexikoa (V2.0) eta urre-patriaren arteko kontigentzia-taula.

orekatua izan da. Hemen, ordea, joera bat nagusitzen da: kasu gehiengotan, batek (urre-patriak) zerrendako hitzari sentimendu-balentzia esleitzen dionean, besteak (lexikoiak) ez du egiten. Horrek lexikoa urre-patriaren aldean kontserbadoreagoa eta zorrotzagoa dela aditzera ematen du. Lexikoiak askozaz ere hitz gutxiagori esleitzen dien balentzia, Pearson 2ko korrelazio koefizientea baxuagoa da.

Laburbilduz, Pearson 1 korrelazio-koefizientea altua eta antzekoa izan da bai bi anotatzaileen artean, bai lexikoiaren eta urre-patriaren artean. Horrek hitzei esleitu zaien sentimendu-balentzia (-5etik +5era) antzekoa (0,79 vs. 0,76) izan dela esan nahi du. Pearson 2 korrelazio koefizientea, aldiz, altua izan da bi anotatzaileen artean, baina jaitsiera bat egon da lexikoiaren eta urre-patriaren artean (0,73 vs 0,54). Lexikoiak urre-patriarekin (eta bi anotzaileekin) alderatuz, hitz gutxiagori esleitu die sentimendu-balentzia eta, horregatik, da koefizientea baxuagoa.

1 Ikerketa hipotesia

1.2 ataleko ikerketa hipotesiari erantzunez, emaitzek erakusten dute posible dela gaztelaniatik euskara sentimenduen lexikoi bat itzultzea eta bera baliagarria izatea.

Pearson korrelazioek erakusten dutenez, itzulpenaren ondoren, urre-patriaren eta sentimenduen lexikoiaren adostasuna 0,76 koefizientekoa hitzei sentimenduen balentzia esleitzekoan. Hala ere, badaude hobetzeko alderdi batzuk, besteak beste, garaturiko sentimenduen lexikoian sarrera gehiago gehitzea falta delako.

4.3 Euskarazko sentimenduen sailkatzailea

Lehenik eta behin, guk sortu dugun sentimenduen sailkatzailearen oinarrian dagoen SO-CAL tresnaren ezaugarriak deskribatuko ditugu (4.3.1 atala). Ondoren, euskarazko sentimenduen sailkatzailea sortzeko SO-CALen egin-dako moldaketak azalduko ditugu (4.3.2 atala). Azkenik, euskarazko sentimenduen sailkatzailearen ebaluazioaren emaitzak aurkeztuko ditugu (4.3.3 atala).

4.3.1. SO-CAL izeneko sentimenduen sailkatzailearen ezaugarriak

Guk sortu dugun euskarazko sentimenduen sailkatzailearen oinarrian SO-CAL tresna (Taboada *et al.*, 2011) dago. Tresna hori hasiera batean ingeleserako sortu bazen ere; gaur egun beste hizkuntzetako bertsiok ere badaude, esaterako txinerarena (Miao *et al.*, 2013).

Tresna horren oinarrian sentimenduen lexikoa dago. Sentimenduen lexikoiak lau gramatika-kategorietako hitzak (izenak, adjektiboak, aditzak eta adberbioak) biltzen ditu. Hitz horiek -5 eta $+5$ arte sentimendu-balentzia dute, 4.16 taulan ikusten den moduan.

Gramatika-kategoria horretako hitzez gain, intentsifikatzaileak ere badaude eta horiek hitzen sentimendu-balentzian aldaketak eragiten dituzte. Intentsifikatzaileen hitzen sentimendu-balentzia $\% +100$ eta $\% -50$ artean indartu edo ahuldu dezakete 4.17 taulan, intentsifikatzaile horietako batzuk ageri dira.

Intentsifikatzaileen eta sentimendu-balentziadun hitzen artean (4.17) adibidean bezalako eragiketako gertatzen dira. *Sleazy* (“zikin”) hitzak -3 sentimendu-balentzia du baina *somewhat* (“zerbait”) intentsifikatzailearen ondorioz, bere sentimendu-balentzia -3 tik -2 , 1era jaisten da.

- (44) Sleazy: (-3) , somewhat sleazy: $-3 \times (\%100 - \%30) = -2,1$. (Taboada *et al.*, 2011, 275 orr.)

SO-CAL tresnak (Taboada *et al.*, 2011) ezeztapena ere tratatzen du eta, ho-

Hitza	Sentimendu-balentzia
monstrosity	-5
hate (izena eta aditza)	-4
disgust	-3
sham	-3
fabricate	-2
delay (izena eta aditza)	-1
determination	+1
inspire	+2
inspiration	+2
endear	+3
relish (aditza)	+4
masterpiece	+5

4.16 taula: SO-CAL tresnako lexikoiaren zenbait sarrera (Taboada *et al.*, 2011).

Intentsifikatzailea	Modifikatzailea (%)
slightly	-50
somewhat	-30
pretty	-10
really	+15
very	+25
extraordinarily	+50
(the) most	+100

4.17 taula: SO-CAL tresnako zenbait intentsifikatzaile (Taboada *et al.*, 2011).

rretarako, 5.2.2 atalean azaldutako desplazamendu-ezeztapena izeneko hurbilpena erabiltzen du. (45) adibidean, tresnak ezeztapena nola laguntzen duen ikus daiteke. Ezeztapenak -4 balentzia du esaldiak orientazio semantiko positiboa duenean eta -4 esaldiak orientazio semantiko negatiboa duenean. (45) adibideko lehen zatian *terrific* (“bikaina”) hizak -5 sentimendu-balentzia du baina ezeztapenak sentimendu-balentzia hori $+1$ era aldatzen du. Adibide bereko bigarren zatian, aldiz, *terrible* (“beldurgarria”) hitzak -5 sentimendu-balentzia du baina ezeztapenak balentzia hori $+1$ ean bihur-

tzen du.

- (45) He's not terrific ($5 - 4 = 1$) but not terrible ($-5 + 4 = -1$) either.
(Taboada *et al.*, 2011, 277 orr.)

Sentimendu lexikoiaren beste ezaugarrietako bat da sentimenduen analisirako baliagarriak ez diren esaldiak identifikatzea eta horiei sentimendu-balentzia ez esleitzea. Sentimenduen analisirako baliagarriak ez diren esaldiei *irrealis* edo testuinguru ez-faktualak deritze. Ingelesean, testuinguru ez-faktuala hitzen hurrenkeraren, aditz modalaren edota aginteraren bidez agertzen da. SO-CAL tresnak (Taboada *et al.*, 2011) horrelako elementuak aurkitzen dituenean, esaldi horietako hitzen sentimendu-balentzia ez du kontuan hartzen. Komatxoaren artean agertzen diren hitzen sentimendu-balentziak ere ez ditu kontuan hartzen.

Esaterako, (46) adibidean, *great* (“sekulako”) hitzak +3 sentimendu-balentzia du baina esaldiak testuinguru ez-faktuala duenez, bertako hitzen sentimendu-balentzia ez aintzat hartzen.

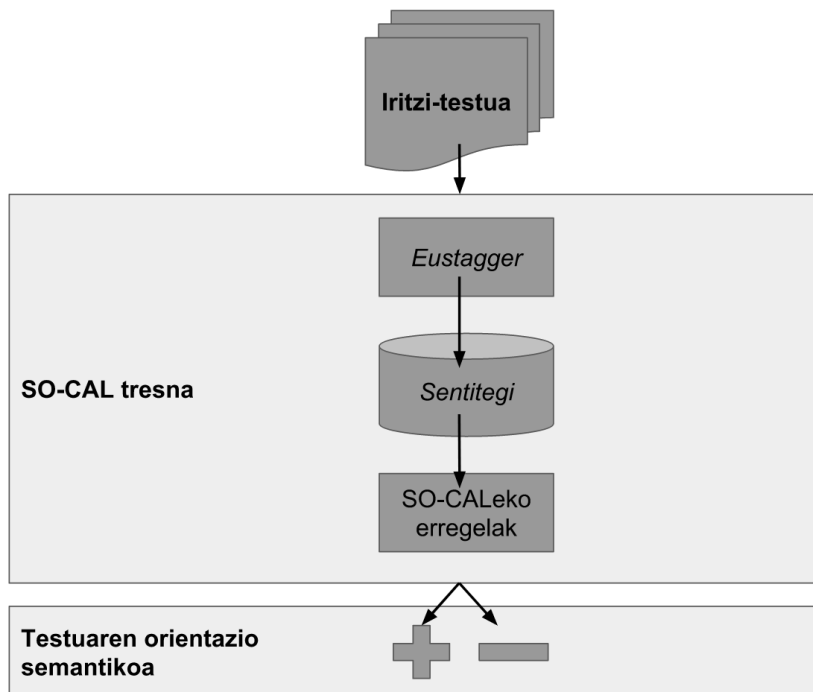
- (46) This should have been a great movie. (+3 = 0). (Taboada *et al.*, 2011, 279 orr.)

Kennedy eta Inkpenek (2006) lexikoian oinarritutako sentimenduen sailkatzailerek orientazio semantiko positiboa esleitzeko joera antzeman zuten eta, horregatik, tresna honetan sentimendu-balentzia negatiboa duten hitzei pisua esleitu zaie. Hain zuzen, % +50eko pisua esleitzen zaie hitz horiei. Gainera, sentimendu-balentziadun hitz bat testu batean hainbat aldiz errepikatzen bada, errepikapen horietako bakoitzean sentimendu-balentziadun hitz horrek bere balentzia galtzen du ($1/n$ eragiketa, n = errepikapen-kopurua). Erregela horren adibidea euskaran 4.7 irudian ikus daiteke.

Azkenik, SO-CAL tresnaren (Taboada *et al.*, 2011) ingelesezko bertsioak zenbait erregela morfologiko ditu, testuetako hitzei flexio-morfemak kendu eta lema moduan uzten dituetan. Hitza lematizatuta dagoenez, tresnako lexikoia hitzari sentimendu-balentzia esleitu diezaioke, baldin eta hitza lexikoian badago.

4.3.2. Sentimenduen sailkatzailearen arkitektura

SO-CAL izeneko sentimenduen sailkatzailearen (Taboada *et al.*, 2011) euskarazko bertsioa garatzeko zenbait aldaketa egin behar izan ditugu. Aldaketa horiek 4.4 irudian daude ikusgai.



4.4 irudia: SO-CAL sentimenduen sailkatzailearen euskarazko bertsioa.

Euskarazko bertsioa sortzeko egin dugun lehen aldaketa *Eustagger* (Aduriz *et al.*, 2003) integratzea izan da. Ingelesak, hizkuntza analitikoa denez, erregela morfologiko gutxi behar ditu flexio-morfemak kendu eta hitzak lematizatuta uzteko. Euskara, ordea, morfologikoki aberatsa da eta, horregatik, testuetako hitzak lematizatzeke *Eustagger* (Aduriz *et al.*, 2003) integratu dugu.

Gainera, ingelesezko sentimenduen lexikoia ordez, guk sortu dugun *Sentitegi* sentimenduen lexikoa (Alkorta *et al.*, 2018) integratu dugu. Gure sentimenduen lexikoiak izen, adjektibo, aditz eta adberbio gramatika-

kategorietakoak biltzen ditu.

Azkenik, SO-CAL tresnak (Taboada *et al.*, 2011) berak dituen bestelako erregelak bere horretan utzi ditugu. Erregela horiek sentimendu-balentzia negatiboa duen hitzei pisua esleitzekoa, testu batean hainbat aldiz errepikatzen diren hitzei balioa kentzekoa edota komatxoaren artean dauden hitzen sentimendu-balentziak kontuan ez hartzekoak dira.

Euskarazko sentimenduen lexikoia garatzeko aldaketa horiek egin ostean, jarraian azalduko dugun moduan ibiltzen da euskarazko sentimenduen lexikoia. 4.5 irudian, euskarazko sentimenduen sailkatzaileak aztertuko duen iritzi-testua ageri da. Ikusten den moduan, testua gordina da.

Asteburu polita orokorrean.
Ez dugu Penintsula hegoaldean dagoen beroa eta ezta nahi ere. Temperatura atseginekin doa uda eta horrelaxe jarraituko dugu asteburuan ere.
Gainera, oro har eguraldia ondo portatuko da. Hala ere, larunbata hodeiekin hasiko dugu eta baliteke zirimiru pixka bat egitea ere, bereziki goizaldean. Arratsalderako ordea, argituko du eta eguzkia ikusiko dugu. Iparraldeko haizea ibiliko da eta temperatura maximoa 21 C-koa izango da, atsegina.
Igandean eguzkia nagusituko da berriro ere, baina udan askotan gertatzen den bezala, goizaldean hodei baxuak sortuko dira seguruenik eta hodei hoiekin esnatzea posible da, baina goizean zehar eta larunbatean baino azkarrago, ostarteak zabaltzen hasiko dira, eguraldi eguzkitsua geratuz eguerditik aurrera. Temperatura pixka bat igo egingo da, maximoa 24 C-ra iritsiz.
Esandakoa gure betiko uda naturalarekin jarraituko dugu: hodei baxuak, eguzkia eta temperatura atseginak, berorik gabe.

4.5 irudia: EGU40 iritzi-testua.

Euskarazko sentimenduen sailkatzaileari 4.5 irudiko iritzi-testua ematean, tresnak lehendabizi testua lematizatzen du, eta ondoren, bertako hitzei sentimendu-balentzia esleitzen die. 4.6 irudian, iritzi-testua lematizatuta dago eta esaldien eskuinetara esaldietako hitzei esleitu dien sentimendu-balentzia agertzen da.

Hurrengo urratsean, sentimendu-balentzia esleitu dien hitzei zenbait erregela aplikatzen dizkie, hala nola sentimendu-balentzia negatiboa duten hitzei pisua esleitzekoa, hainbat aldiz errepikatzen diren hitzei balioa kentzekoa eta komatxoaren artean dauden sentimendu-balentziadun hitzak kontuan ez hartzekoa.

```

asteburu polit orokor . 7.0
tenperatura atsegin joan uda eta horrelaxe jarraitu *edun asteburu ere . 3.3
gain , oro har eguraldi ondo portatu izan . -1.5
hala ere , larunbat hodei hasi *edun eta *edin zirimiri pixka bat egin ere ,
bereziki goizalde . 0
arratsalde ordea , argitu *edun eta eguzki ikusi *edun . 2.0
iparraldeko haize ibili izan eta tenperatura maximo 21 C-koa izan izan ,
atsegin . 4.0
igande eguzki nagusitu izan berriro ere , baina uda asko gertatu izan bezala
, goizalde hodei baxu sortu izan seguru eta hodei hoiekin esnatu
posible izan , baina goiz zehar eta larunbat baino azkar , ostarte
zabaldtu hasi izan , eguraldi eguzkitsu geratu eguerdi aurrera . 5.8
tenperatura pixka bat igo egin izan , maximo 24 C-ra iritsi . 0.5
esan gu be uda natural jarraitu *edun : hodei baxu , eguzki eta tenperatu
atsegin , bero gabe . 1.75

```

4.6 irudia: EGU40 iritzi-testua lematizatuta eta hitzei sentimendu-balentziak esleituta.

4.7 irudian, euskarazko sentimenduen sailkatzaileak iritzi-testuko sentimendu-balentziadun hitzei aplikatutak erregelak ikus daitezke. Ikusten denez, zenbait hitzek sentimendu-balentzia negatiboa dute eta horiei pisua esleitu die (NEGATIVE erregela). Ondorioz, esaterako, adjektiboetan *baxu* hitzaren sentimendu-balentzia -3tik -4,5era pasa da. Beste erregela batek, testu berean hainbat aldiz errepikatzen diren hitzei balioa kentzen die. *Atsegin* horren adibidea da. Lehen aldiz agertu denez, tresnak *atsegin* hitzari +1 sentimendu-balentzia esleitu dio, baina bigarren aldiz agertu denean, hitzari pisua kendu dio (REPEATED erregela) eta bere sentimendu-balentzia +1etik +0,5era igaro da.

Sentimendu-balentziadun hitzei erregela horiek aplikatu ostean, tresnak iritzi-testuari sentimendu-balentzia esleitzen dio, 4.8 irudian ikusten den moduan. Kasu honetan, iritzi-testuaren sentimendu-balentzia +1,04 da; beraz, iritzi-testuak balorazio positiboa egiten du.

4.3.3. Sentimenduen sailkatzailearen ebaluazioa

Atal honetan, sentimenduen sailkatzailearen ebaluazioaren emaitzen berri emango dugu. Ebaluazioaren emaitzak hainbat modutara emango ditugu: asmatutako eta ez asmatutako testuen orientazio semantikoak, orientazio semantiko

```

Text Length: 168
-----
Nouns:
-----
har -1.0 X 1.5 (NEGATIVE) = -1.5
-----
Average SO: -1.5
-----
Verbs:
-----
jarraitu *edun asteburu ere 1.0 X 1.3 (INTENSIFIED) = 1.3
argitu 2.0 = 2.0
ibili 2.0 = 2.0
atsegin 1.0 = 1.0
nagusitu izan berriro ere 1.0 X 1.3 (INTENSIFIED) = 1.3
sortu 2.0 = 2.0
jarraitu 1.0 X 1/2 (REPEATED) = 0.5
atsegin 1.0 X 1/2 (REPEATED) = 0.5
-----
Average SO: 1.325
-----
Adjectives:
-----
polit 4.0 = 4.0
orokor 3.0 = 3.0
atsegin 2.0 = 2.0
maximo 1.0 = 1.0
baxu -3.0 X 1.5 (NEGATIVE) = -4.5
seguru 1.0 = 1.0
posible 1.0 = 1.0
azkar 2.0 = 2.0
eguzkitsu 3.0 = 3.0
maximo 1.0 X 1/2 (REPEATED) = 0.5
natural 2.0 = 2.0
baxu -3.0 X 1/2 (REPEATED) X 1.5 (NEGATIVE) = -2.25
bero 1.0 = 1.0
-----
Average SO: 1.05769230769
-----

```

4.7 irudia: EGU40 iritzi-testuko hitzei sentimenduen sailkatzailearen erregelak aplikatuta.

```

-----
Total SO: 1.03863636364
-----

```

4.8 irudia: EGU40 iritzi-testuaren sentimendu-balentzia.

jakin bakoitzean sailkatzaileak duen asmatze-tasa eta, azkenik, sailkatzaileak corpuseko domeinu bakoitzean duen asmatze-tasa.

4.18 Taulako emaitzen arabera, sailkatzailearen asmatze-tasa 0,71koa da. Izan ere, corpuseko test zatian dauden 48 testuetatik 34 iritzi-testuren orientazioa semantikoa ondo esleitu baitu. Iritzi-testuek duten orientazio semantikoaren ikuspegitik emaitzak aztertuta, emaitzek bestelako interpretazioak egiteko aukera ematen dute.

Emaitzak	Testu-kopurua	%
Ondo	34	0,71
Gaizki	14	0,29

4.18 taula: Sentimenduen sailkatzailearen ebaluazioaren emaitzak.

4.19 taulako emaitzetan, tresnaren funtzionamenduari buruzko emaitza argigarriak ageri dira. Ikusten den moduan, sentimenduen sailkatzailearen asmatze-tasa orientazio positiboko iritzi-testuetan lekoa da. Emaitzak zeharo ezberdinak dira orientazio negatiboko iritzi-testuen kasuan. Mota horretako iritzi-testuekin, tresnaren asmatze-tasa 0,33koa da; izan ere, 24 orientazio negatiboko iritzi-testu egon arren, 8 iritzi-testu bakarrik identifikatu ditu negatibo moduan. Beraz, esan liteke, sentimenduen sailkatzaileak joera bat duela iritzi-testuei orientazio semantiko positiboa esleitzeko. Joera hori bat dator Alistair eta Dianak (2005) dioenarekin. Diotenez, lexikoian oinarritutako sentimenduen sailkatzaileek orientazio semantikoa esleitzeko joera erakusten du eta hori lotuta dago unibertsalki gizakiek hizkuntza positiboa erabiltzeko duten ohiturarekin Boucher eta Osgood (1969).

Emaitzak	Asmatutako testu-kopurua	%
Positiboak	24/24	1
Negatiboak	24/8	0,33

4.19 taula: Sentimenduen sailkatzailearen ebaluazioaren emaitzak iritzi-testuen orientazio semantikoa aintzat hartuta.

Azkenik, emaitzak domeinuen arabera aztertuta, domeinu guztietan asmatze-tasa antzekoa dela ikus daiteke, literaturan izan ezik.

4.20 taulak erakusten duenez, domeinuetan asmatze-tasa txikiena eguraldian, musikan eta politikan dago; asmatze-tasa 0,63koa da. Kirola eta zinema tartean kokatzen dira 0,75eko asmatze-tasarekin eta, azkenik, emaitza onenak literaturak lortzen ditu, testu baten orientazio semantikoa izan ezik, beste guztiena asmatu baitu tresnak bertan (0,88ko asmatze-tasa). Domeinutik domeinura ez dago ezberdintasun handirik, domeinu bakoitzean orientazio semantiko positibo eta negatiboko iritzi-testu kopuru bera dagoelako.

Emaitzak	Testu-kopurua	%
Eguraldia	8/5	0,63
Musika	8/5	0,63
Politika	8/5	0,63
Kirola	8/6	0,75
Zinema	8/6	0,75
Literatura	8/7	0,88

4.20 taula: Sentimenduen sailkatzailearen ebaluazioaren emaitzak domeinuen arabera.

4.4 Laburpena

Kapitulu honetan, tesi-lanean garatutako baliabideak zein diren azaldu dugu. Lehenik eta behin, Euskarazko Iritzi Corpora izeneko sei domeinutako 240 iritzi-testu biltzen dituen corpora (Alkorta *et al.*, 2016) eratu dugula aipatu behar da.

Bigarrenik, sortu dugun *Sentitegi* izeneko euskarazko sentimenduen lexikoia-
ren (Alkorta *et al.*, 2018) berri ere eman dugu. Ikusi den moduan, sentimen-
duen lexikoia-
ren bi bertsio sortu ditugu, bata domeinuari lotu gabea (8.140 sarrerakoa) eta bestea, corpuseko domeinuei lotutakoa (1.237) eta izaera mu-
rriztaileagoa duena, kolokazioak eta adierazpenak ez baititu kontuan hartzen. Sorkuntzak emandakoa esplikatu ondoren, ebaluatu egin dugu. Horretarako, Euskarazko Iritzi Corpuseko maiztasun handieneko lehen 100 hitzak hartu ditugu, bi anotatzaileek hitzei sentimendu-balentzia esleitu die eta ondoren, bertatik urre-patroia sortu da. Jarraian, urre-patroi hori euskarazko senti-
menduen lexikoia-
k zerrendako hitzei eman dien sentimendu-balentziarekin alderatu dugu. Emaitzek erakusten dutenez, sentimenduen lexikoia-
k hitzek duten sentimendu-balentzia egokia da (Pearson korrelazioa 0,76koa delako), baina berez sentimendu-balentzia eduki beharko luketen hitz batzuei lexi-
koiak ez die balentzia esleitu eta, ondorioz, Pearson korrelazioa 0,54koa da.

Azkenik, tesi-lan honen hirugarren emaitza euskarazko sentimenduen sailka-
tzailea da. Euskarazkoa sortzeko ingelesezko SO-CAL sentimenduen sailka-
tzailean oinarritu gara eta zenbait aldaketa egin behar izan ditugu. Esatera-
ko, *Eustagger* testu lematizatzailea (Aduriz *et al.*, 2003) inplementatu dugu, baita *Sentitegi* euskarazko sentimenduen lexikoia (Alkorta *et al.*, 2018) ere. Sentimenduen sailkatzaile hori ebaluatu egin dugu eta, emaitzek erakusten dutenez, tresnak kasuen % 71etan iritzi-testuaren orientazio semantikoa ondo sailkatzen du.

**BALENTZIA
ALDATZAILEAK
HIZKUNTZA MAILA
EZBERDINETAN**

Balentzia-aldatzaileak

Kapitulu honetan, hizkuntzako maila ezberdinetan dauden euskarazko balentzia-aldatzaileak identifikatzeko lanaren emaitzak azalduko ditugu. Guztira lau hizkuntza maila landu ditugu: fonologiko eta morfologikoa (5.1 atala), sintaktikoa (5.2 atala) eta diskurtsokoa (5.3 atala).

5.1 Maila fonologiko eta morfologikoa

5.1.1 atalean, balentzia-aldatzaile fonologiko eta morfologikoen sailkapena hiru modutan aurkeztuko dugu. 5.1.2 atalean, hainbat balentzia-aldatzaile morfologiko dituzten hitzei buruz arituko gara, 5.1.3 atalean, eragin ezberdinetako balentzia-aldatzaileen eta orientazio semantiko ezberdinetako hitzen arteko konbinazioak azalduko ditugu. Azkenik, 5.1.4 atalean, maila fonologikoak eta morfologikoak euskaran duten garrantzia nabarmenduko dugu.

5.1.1. Balentzia-aldatzaileen sailkapena

5.1 taulan, Euskarazko Iritzi Corpusetik (Alkorta *et al.*, 2016) lortutako balentzia-aldatzaile fonologiko eta morfologikoen zerrenda ageri da maiztasunean oinarrituz sailkatuta. Guztira 32 balentzia-aldatzaile aurkitu ditugu eta horiek guztiak 1.623 aldiz agertzen dira corpusean. Gehienak atzizkiak dira eta hiru bustidura adierazkor eta aurrizki mota ageri dira.

Balentzia-aldatzaileek sentimendu-balentzian duten eraginean oinarritutako sailkapena 5.2 taulan dago. Bertan, balentzia-aldatzaileak lau multzotan

Hizkia	Maiztasuna	Adibidea
-tasun	307	zailtasun
-garri	250	barregarri
-tsu	170	gatazkatsu
-en	163	handien
-gabe	71	akatsgabe
des-	62	desorekatu
-egi	61	bikainegi
-dun	60	berezidun
-gile	60	inbertsiogile
-keria	57	handikeria
-s-/-z- >-x-	39	goxo
-ez/ez-	39/0	ezezonkor
-kide	38	garaikide
-ezin/ezin-	36/3	garaiezin
-xe/-txe	30	oraintxe
-txo	27	erasotxo
-ezia	27	dotorezia
-zale	27	bizizale
-gintza	16	osasungintza
-t- >-tt-	12	puntu
-gai	12	osagai
-gaitz	10	ulergaitz
-gura	8	logura
-gailu	8	neurgailu
-min	7	ikusmin
-ska/-xka	6	herrixka
-z- >-tx-	6	txoratuta
-tza	4	burujabetza
-tzar	3	krisitzar
-nahi	2	handinahi
-txa	1	neskatxa
a-	1	atenporalitatea
Guztira	1.623	

5.1 taula: Euskarazko Iritzi Corpusetik (Alkorta *et al.*, 2016) lortutako balentzia-aldatzaile morfologiko eta fonologikoak.

banatu ditugu: i) balentzia indartzen dutenak, ii) ahultzen dutenak eta iii) balentzian eraginik ez dutenak.

Indartu	Ahuldu	Eraginik ez
-garri	-gabe	-ezia
-tsu	des-	-tasun
-en	-egi	
-s- / -z- > -x-	-keria	
-zale	-ez / ez-	
-t- > -tt-	-ezin / ezin-	
-gura	-xe / -txe	
-min	-txo	
-z- > -tx-	-gaitz	
-tza	-ska / -xka	
-tzar	-txa	
-nahi	a-	

5.2 taula: Balentzia-aldatzaile morfologiko eta fonologikoak sentimendu-balentzian duten eraginaren arabera sailkatuta.

Bukatzeko, 5.3 taulan, Euskarazko Iritzi Corpusetik (Alkorta *et al.*, 2016) lortutako hizkiak duten balio semantikoa irizpidetzat hartuta sailkatuta ageri dira. Bertan, hizkiak eta bustidura adierazkorra 12 multzo semantikotan banatu ditugu. Lehenik eta behin, ezeztapenarekin lotutakoak daude. Ezeztapenezko hizki horiek ezeztapen-marka moduan antzeko modu batean eragiten diote sentimendu-balentziari, baina hitzari eragin beharrean, esaldi edo sintagmari eragiten diote. Multzo horretan aurrizkiak eta atzizkiak aurki daitezke.

Datorren multzoan konparazioarekin lotutako hizkiak ageri dira: *-egi* hizkia, sentimendu-balentzia ahultzen duena eta *-en* hizkia, sentimendu-balentzia indartzen duena.

Hurrengo bi multzoak tamainarekin lotutakoa dira. Semantikoki handigarri funtzioa betetzen duen balentzia-aldatzaile bakarra aurkitu dugu: *-tzar*. Txikigarri funtzioa duten balentzia-aldatzaileak, aldiz, asko dira eta denak atzizkiak dira. Multzoa esanahiarekin lotura duten bi balentzia-aldatzaile aurkitu ditugu: *-tsu* eta *-tza* eta biek sentimendu-balentzia indartzen dute.

Ezeztapena	Konparazioa	Handigarria	Txikigarria
des-	-egi	-tzar	-xe/-txe
-ez/ez-	-en		-txo
-ezin/ezin-			-ska/-xka
-gabe*			-txa
(a-)			
Zailtasuna	Nolakotasuna	Lanbidea	Jabetza
-gaitz	-tasun	-gile	-dun
-garri	-keria	-gintza	-gabe*
	-ezia		
Nahia	Hurbiltasuna/samurtasuna	Multzoa	Bestelakoak
-gura	-s-/-z- >-x-	-tza	-zale
-nahi	-t- >-tt-	-tsu	
-min	-z- >-tx-		

5.3 taula: Euskarazko Iritzi Corpusetik lortutako hizkien sailkapen semantikoa.

Zailtasunarekin lotura duten beste bi balentzia-aldatzaile ere badaude: *-gaitz* eta *-garri*. Lehenak sentimendu-balentzia ahultzen du eta bigarrenak, berriz, sentimendu-balentzia indartzen du.

Bestalde, nolakotasuna adierazten duten hiru balentzia-aldatzaile daude, baina, *-keria* hizkiak sentimendu-balentzia ahultzen du eta *-tasun* eta *-ezia* atzizkiek sentimendu-balentzia indartzen dute.

Jabetza adierazten duten bi atzizki daude: *-dun* eta *-gabe*. Bigarren atzizkiak semantikoki jabetza edo ezeztapena adieraz dezake eta, ondorioz, bi multzoetan sailkatu dugu. Azken honek sentimendu-balentzia ahultzen du. Bustidura adierazkorrei dagokienez, hirurek semantikoki hurbilpena edota samurtasuna adierazten dute eta sentimendu-balentzia indartzen dute.

Azkenik, aipaturiko multzoekin zerikusirik ez duten balentzia-aldatzaile bakararra dago. Hizkia *-zale* da. Zailtasuna adierazten du eta sentimendu-balentzia indartzen du.

5.1.2. Hainbat balentzia-aldatzaile morfologiko dituzten hitzak

Hitz batean hainbat hitz agertzen direnean, hitz horrek egitura jakin bat du. Bi ezaugarri dituzte mota horretako hitzek:

- Hainbat balentzia-aldatzaile morfologiko eta balentzia-aldatzaile fonologikoa den bustidura adierazkorra batera agertzen diren hitzik ez dago. Hau da, hitz batean hainbat balentzia-aldatzaile morfologiko badaude, bustidura adierazkorrik ez da agertzen.
- Hitz batean gehienez hiru balentzia-aldatzaile morfologiko ager daitezke batera bata bestearen atzetik. Eta bitartean, hitzaren sentimendu-balentzian eta orientazio semantikoak aldaketak gertatzen dira.

5.4 taulan, balentzia-aldatzaile morfologiko bat baino gehiago dituzten hitzen egituraketa ageri da. Guztira, mota horretako 36 hitz eta 11 hizki aurkitu ditugu. Gehienez hiru hizki ager daitezke elkarren segidan eta hizkiak erantsi ahala hitzaren sentimendu-balentziak aldaketa ezberdinak jasaten ditu. Hirugarren mailara iristeko ezinbestekoa den bigarren mailan *-en* hizkia egoitea. Eta bigarren mailatik hirugarrenera balentziaren ahultzea gertatzen da, nahiz eta hasieratik egoeratik balentzian indartze bat dagoen.

Hizkiak konbinatzeko aukerak eta beren ondorioak askotarikoak dira. Posible da hizkiak gehitu ahala hitzaren sentimendu-balentzia indartzea. Hori gertatzen da *prestigiotsuen* hitzean, *-tsu* eta *-en* hizkiek hitzaren balentziaren positibotasuna indartzen baitute. Hizkiak gehitu ahala hitzaren sentimendu-balentzia negatiboago bilakatzen ere joan daiteke. *Lotsagabekeria* hitzean, *-gabe* eta *-keria* hizkiek *lotsa* hitzaren sentimendu-balentzia ahuldu eta negatibo bilakatzen dute. Tarteko zerbait ere gerta daiteke, hots, hitzaren balentzia hasieran indartzea eta gero ahultzea. *Berritsukeria* hitzaren kasuan, *-tsu* hizkiak hitza positiboago bilakatzen du eta ondoren, *-keria* hizkiak positibotasun hori negatibo bilakatzen du.

5.4 taulan eta azalpenean adibide moduan jarritako hitzak (adibidez, *berritsu* eta *berritsukeria*) *Elhuyar* (Elhuyar, 2013) eta *Zehazki* (Sarasola, 2005) hiztegietan egon daitezke. Hiztegietak sarrerak direnez, sentimenduen lexikoian

Maila	Balentzia	Hizkia
0	handi gehi harritu malenkonია aspertu berri prestigio lotsa garrantzi axola garbiz zehatz	
1	handinahi gehiegi harrigarri malenkoniatsu aspergarri berritsu prestigiotsu lotsagabe garrantzitsu axolagabe garbizale zehazgabe	-nahi (+) -egi (-) -garri (0) -tsu (+) -gabe (-) -zale (+)
2	handinahikeria gehiegikeria harrigarrien malenkoniatsuago aspergarriago berritsukeria prestigiotsuen lotsagabekeria garrantzitsuen axolagabeen garbizaletasun zehazgabetasun	-en (+) -tasun (0) -keria (-) -ago (+)
3	harrigarrienetako prestigiotsuenetariko garrantzitsuenetako	-eneta(ri)ko (+)

5.4 taula: Hitz batean hainbat hizki agertzeko moduak eta horien eragina hitzaren balentzian.

ager daitezke. Sentimenduen balentzian koherentzia mantentzeko, beharrezkoa da *berritsu* hitzak duen sentimenduen balentziari *-keria* balentzia-aldatzailearen eragina aplikatuta, eragiketa horrek eta sentimenduen lexikoian legokeen *berritsukeria* hitzak sentimendu-balentzia bera edukitzea.

5.1.3. Balentzia-aldatzaileak orientazio semantiko ezberdinetako hitzetan

5.1.1 atalean, balentzia-aldatzaile fonologiko eta morfologikoen balentzia indartu, ahuldu edo balentzian eraginik ez dutela adierazi dugu. Eragin ezberdinetako balentzia-aldatzaile horiek orientazio semantiko ezberdinetako hitzekin batu daitezke eta, ondorioz, hitz horien sentimendu-balentzian gertatzen diren aldaketak askotariakoak dira.

5.5 taulan konbinazio posible guztiak ikus daitezke. Aintzat hartuta hitz batek ++ edo -- orientazio semantikoa duela, balentzia horiek aldaketa ezberdinak jasan ditzakete balentzia-aldatzaileen ezaugarrien arabera. Are positiboago edo negatiboago bilaka daitezke edo, bestela, negatibo edo positibo izaterantz jotzen dute, kasu batzuetan, zeinu aldaketa jasateraino.

Hizkia / hitza	Hizkia (↑)	Hizkia (↓)	Hizkia (0)
Hitza (+)	+++	+/-	++
Hitza (-)	---	-/+	--
Hitza (0)			

5.5 taula: Balentzia aldatzen duten hizkiek sentimendu-balentzia ezberdineko hitzetan duten eragina aztertzeke taula.

Ondoren, indartzaileak, ahultzaileak edota eraginik ez duten balentzia-aldatzaileak orientazio semantiko ezberdinetako hitzekin batzean gertatzen diren aldaketak aurkeztuko ditugu.

- Sentimendu-balentzia positibodun hitz bat eta balentzia indartzailea den hizki bat.
Batura honetan, hitzaren balentzia positiboa are positiboago bilakatzen da. Esaterako, *eroso* hitzak +2 balentzia du eta *erosoen* hitzak balentzia are positiboagoa du.

- Sentimendu-balentzia positibodun hitz bat eta balentzia ahultzailea den hizki bat.
Baturaren ondorioz, hitzak balentzia positiboa manten dezake, nahiz eta ahulagoa izan, edota hitzaren balentzia positiboa negatibo bilaka daiteke. *Ondu* aditzak +2 balentzia du eta *des-* ezeztapenezko balentzia ahutzailearekin batuz gero, balentzia negatibo bilakatzen da. Emaidza *desondu* -2 da¹. *Estilistiko* (+3) hitzari *-egi* hizkia gehituz gero, aldiz, *estilistikoegi* hitzaren balentzia jaisten da.
- Balentzia positibodun hitz bat eta balentzia-aldatzailea ez den hizki bat.
Konbinaketa honetan, hitzak bere balentzia mantentzen du. *Lasai* adjektiboak +2 balentzia du eta *lasaitasun* hitzak ere sentimendu-balentzia bera mantentzen du.
- Hitzak sentimendu-balentzia negatiboa eta balentzia indartzaile bat.
Horren ondorioa hitzaren sentimendu-balentzia are negatiboago bilakatzear da. *Kalte* izenak -2 balentzia du *-tzar* hizki handigarria erabilita, *kaltetzar* hitzak are balentzia negatiboagoa du.
- Balentzia negatibodun hitz bat eta balentzia ahultzailea den hizki bat.
Kasu horretan, hitzaren balentziak negatibo izaten jarraitzen du, baina intentsitatea ahulago batez edo, bestela, hitzaren balentzia negatiboa positibo ere bilakatu daiteke. *Sentitegi* lexikoian (Alkorta *et al.*, 2018), *labur* hitzak -1 balentzia du eta *laburregi* hitzak balentzia are negatiboa izango luke. Baina, *nahasi* hitzari, -2 balentziarekin, *-ezin* atzizkia erantsiz gero, *nahastezin* sortzen da eta *-ezin* ezeztapena denez eta ezeztapenak ± 4 balentzia duenez, hitza ahuldu eta +2 sentimendu-balentzia izatera igarotzen da.
- Balentzia negatibodun hitz bat eta balentzian eragiten ez duen hizki bat.
Baturaren ondorioz, hitzaren sentimendu-balentzian ez da aldaketarik

¹Lan honetan, ezeztapena lantzeko hautatu dugun hurbilpena desplazamendu-ezeztapena da eta, horregatik, ezeztapen-markei nahiz ezeztapenezko hizkiei ± 4 balentzia esleitu diegu.

gertatzen. *Baldar* nahiz *baldartasun* hitzek -2 sentimendu-balentzia dute, *-tasun* hizkiak ez duelako balentzian eraginik.

- Balentziarik gabeko hitz bat eta balentzia indartzailea den hizki bat. Hitzak balentziari ez duenez, balentzia-aldatzaileak ez du eraginik. Esaterako, *kolore* eta *mendi* hitzak neutralak dira *Sentitegi* sentimenduen lexikoiaren (Alkorta *et al.*, 2018) arabera, hau da, ez dute balentziarik, eta *koloretsu* eta *menditzar* hitzek ere ez dute.
- Balentziarik gabeko hitz bat eta balentzia ahultzailea den hizki bat. Kasu honetan ere, hitzak sentimendu-balentziarik ez duenez, ez dago balentziaren aldaketarik. Adibidez, *mendi* hitza eta *-txo* balentzia ahultzailea den hizkia batzean (*menditxo*), hitzaren sentimendu-balentzian ez da aldaketarik gertatzen.
- Balentziarik gabeko hitz bat eta balentzia-aldatzailea ez den hizki bat. Konbinazio honetan ere, hitzaren sentimendu-balentzian ez da aldaketarik gertatzen. Horren adibide den *koloretasun* hitzak (*kolore* + *-tasun*) ez du sentimendu-balentziarik eta balentzia-aldaketarik ere jasaten.

5.1.4. Maila fonologikoa eta morfologikoaren garrantzia euskaran

Polanyi eta Zaenenek (2006) ez dute balentzia-aldatzaile morfologiko eta fonologikoen inguruko aipamenik egiten testuinguruko balentzia-aldatzaileen azalpenean.

Baina, lan horrek balentzia-aldatzaileez ari denean, oro har, ingelesa hartzen du ikergaitzat. Horregatik, bertan zerrendatzen diren balentzia-aldatzaile mota guztiak beste hizkuntzetan ezin daitezke baliagarriak izan edota beste hizkuntzetan ingelesean ez dauden balentzia-aldatzaileak egon litezke.

Gure ustez, hori gertatzen da fonologian eta, batez ere, morfologian egon daitezkeen balentzia-aldatzaileekin. Euskara eta ingelesa morfologia tipologiaren aldetik ezberdinak dira, Ingelesa hizkuntza analitikoa da (Szmrecsanyi,

2012), hots, esaldietan hitzen arteko loturak hitz laguntzaileen bidez (preposizioak eta partikulak, esaterako) edo hitz hurrenkeraren bidez egiten dira flexioarekin egin beharrean. Euskara, aldiz, hizkuntza sintetiko eranskaria da (Euskaltzaindia, 1985). Morfologiak garrantzi handia du eta eranskaritasuna erabiltzen du morfemak erabiltzeko. Morfemok hitz erroaren esanahian aldaketak sor ditzakete. Har ditzagun bi adibide:

- (47) Hori ikusi zuenean, irribartxoa atera zitzaion aurpegian.
(48) When she saw that, a giggle came on his face.

Bi adibide horietan, euskarazko eta ingelesezko adibide bana dugu eta esanahi bera dute, baina esanahi bera adierazteko moduak ezberdinak dira. (47) adibidean, irribarrea txikia dela adierazteko, *-txo* hizkia erabiltzen da. Ingelesezko (48) adibidean, ordea, beste estrategia bat erabiltzen da eta irribarre txikia adierazteko *giggle* hitza erabiltzen da. Arrazoi horrengatik eta euskararen morfologia oso aberatsa delako, morfologian egon daitezkeen balentzia-aldatzaileak aztertzea erabaki dugu.

Bestalde, fonologiako balentzia-aldatzaileak aztertzearen arrazoia beste bat da. Balentzia-aldatzaileen azterketari ikuspegi orokor bat eman nahi diogu eta, horrengatik, fonologia ere lanean sartu nahi dugu. Harluxet hiztegiak (Fundazioa, 1995) dionez, bustidura fonema ez-sabaikari bat testuinguru jakin batean sabaikari bilakatzean datzan fenomeno. Bustidura adierazkorra bereiztu egin behar da asimilazio bustiduratik. Oñederrak (1990) horrela egiten du bi bustiduren arteko bereizkuntza:

Euskarazko fonologiaz ari garela, bustikuntza bi motatakoa izan daitekeela izan behar da kontuan: katea fonikoan ingurune jakin batean asimilazioz sortua edo, labur adierazteagatik, bustidura adierazgarria esaten zaiona. (Oñederra, 1990, 13 orr.)

Gaineratzen duenez, bi bustidurak hots klase berari jartzen zaizkio eta aldaketaren ondorioa ere bietan berdina da fonetikan gertatzen den aldaketa berdina delako. Berdintasunak aipatzeaz gain, desberdintasunak zertan diren ere aipatzen du:

Bata ala bestea gertatzeko baldintzak dira aldatzen direnak, asimilazioa katea fonikoko inguruneak sortutako bilakabide sintagmatikoa den bitartean, bustidura adierazgarria hizkuntzaren eremu semantikoari (*lato sensu*) bait dagokio. (Oñederra, 1990, 14 orr.)

Hots, asimilazio bustiduran, asimilazioa gertatzearen baldintza fonetikoa da eta bustidura adierazkorrean, semantikoa. Beste modu batean esanda, hitz batean aldaketa semantikoa gauzatzeko egiten da bustidura adierazkorra eta nahita eragindako aldaketa fonetikoa da. Bustidura adierazkorraz txikitasuna, maitasuna edo goxotasuna adierazi nahi da.

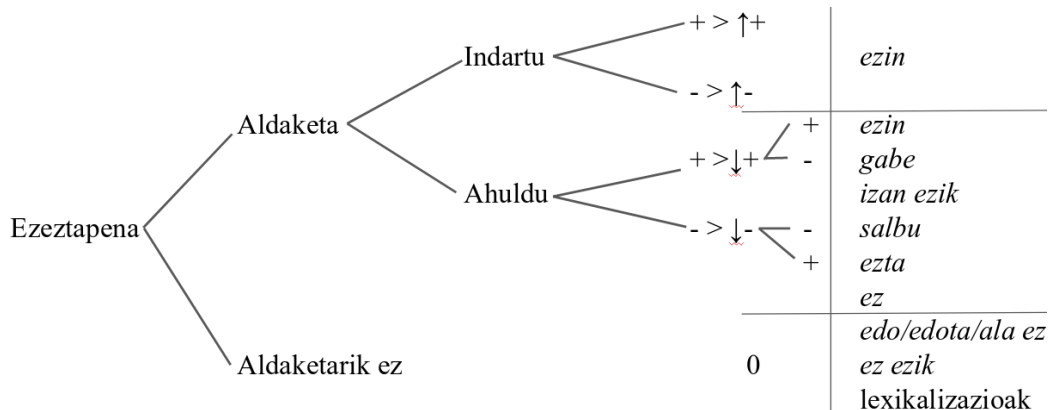
Bustidura adierazkorrek hitzen aldaketa semantikoa eragiten duenez, sentimenduen analisiaren ikuspegitik, balentzia-aldatzaile bat da eta horretan datza fenomeno fonologiko horren interesak.

5.2 Sintaxi maila

Sintaxi maileko 5.2.1 azpiatalean, ezeztapenezko balentzia-aldatzaileak aurkeztuko ditugu. 5.2.2 azpiatalean, trukaketa- eta desplazamendu-ezeztapenaren arteko ezberdintasunak azalduko ditugu, eta desplazamendu-ezeztapena metodologian zergatik erabili dugun ere adieraziko dugu. Azkenik, 5.2.3 azpiatalean, ezeztapen-markak eta beren irismena identifikatzeko erregelak *Murriztapen Gramatikan* (Karlsson *et al.*, 1995) nola sortu eta ondoren, nola ebaluatu ditugun azalduko dugu.

5.2.1. Ezeztapenezko balentzia-aldatzaileak

5.1 irudian, ezeztapenezko balentzia-aldatzaileak sentimendu-balentzian nola eragiten dute ikus daiteke.



5.1 irudia: Ezeztapen-marken eragin ezberdinak sentimendu-balentzian.

Ezin sentimendu-balentzia indartu dezakeen ezeztapen-marka bakarra da. Bestalde, sentimendu-balentzia ahultzen duten ezeztapen-markak gehiago dira: *ezin*, *gabe*, *izan ezik*, *salbu*, *ezta* eta *ez*. Azkenik, badaude zenbait egitura non ezeztapen-marka agertzen den eta ez duen sentimendu-balentzian eragiten. Egitura horiek ezeztapena juntagailuarekin, kontrastezko ezeztapena eta lexikalizatutako egiturak dira. Jarraian, 5.1 irudian azalduko sailkapena xeheguago emango dugu.

Ezeztapen-markaren instantzia guztietatik (359), zazpi kasu antzeman ditugu ezeztapen-markak irismenaren sentimendu-balentzia indartu duena. Balentzia-aldaketa hori *ezin* ezeztapen-markak adjektiboari edo aditzondoari lotuta dagoenean bakarrik gertatzen da (corpusean agertu diren kasu guztien % 1,96).

(49) Dena nahasten da maisulan ezin ederragoa₊₄ osatzeko. (MUS21).²

(49) adibidean, sentimendu-balentzia indartzen den adibide bat dago. Kasu horretan, ezeztapen-markak (*ezin*) konparaziozko atzizkia duen adjektiboari eragin eta bere sentimendu-balentzia indartu du. Hots, positiboa zen sentimendu-balentzia are positiboagoa bihurtzen du. Izan ere, esaldi horretan *ederragorik* ez dagoela esatea *ederrena* dela esatea da; beraz, sentimendu-balentziaren indartze bat dago.

Kasu gehienetan, aztertu dugun corpuseko instantziek sentimendu-balentzia ahultzen dute. Hainbat ezeztapen-marka motek ahultzen dute sentimendu-balentzia: i) *ez*, ii) *gabe*, iii) *ezin*, iv) *izan ezik*, v) *salbu* eta azkenik, vi) *ezta*. Kasuen % 89,98etan (323 instantzietan) gertatu da sentimendu-balentziaren ahultzea.

(50) Horrek ez die eragotzi₋₂ ordea, 57 milioi euro ematea
San Mames klub pribatuari! (POL30)

(51) Bruch-en kontzertua, munduko interesgarriena izan gabe, lan erakargarria da, oso, erraza entzuteko, eta bakarlari honen bertsioak dotoreziaren *plusa* izan zuen dudarik gabe. (MUS22)

(52) Irabazi₊₂ ezinik jarraitzen du Eibarrek, baina oso puntu ona eskuratu du Getaferen zelaian. (KIR17)

(53) “Ez baitute ezertarako balio izan, egun bateko soldata galtzeko₋₂ izan ezik”. (POL17)

²Esaldian dauden anotazio ezberdinen esanahiak hauek dira: **beltz koloreak** ezeztapen-marka adierazten du, azpimarkatutakoak ezeztapenaren irismen-eremua seinatzen du. Berez, ezeztapen-markek ez lukete egon beharko azpimarkatuta, baina hori ez *ezin* ezeztapen-markaren kasuan. Izan ere, ezeztapen-marka horrek bi esanahi ditu: i) ezeztapena bera eta ii) posibilitatea. Horregatik, azpimarkatuta dago.

- (54) E.T. eta Indiana Jones filmak salbu, noski. (ZIN18)
- (55) Ezta Euskal Herria, Espainiako Estatua, Portugal edo Italia moduko beste herrialdeei ere. (POL08)

(50) adibidean, ezeztapen-markaren eragin-eremua esaldi osoa da eta *eragotzi* aditzaren -2 sentimendu-balentzia ahultzen du. Beraz, zeinu positiboa ($+4$) hartzen du ezeztapen-markaren eraginez. (51) adibidean, *gabe* ezeztapen-markak sintagma bati eragiten dio (*munduko interesgarriena izan*). (52) adibidean, *ezin* ezeztapen-markak bere ezkerretara dagoen *irabazi* aditzaren $+2$ sentimendu-balentzia ahultzen du. (53) adibidean, ezeztapen-marka *ezik* da eta *izan* aditzarekin batera *egun bateko soldata galtzeko* sintagma ezeztatzen du. (54) adibidean, *salbu* ezeztapen-markak izen-sintagma bati (*E.T. eta Indiana Jones filmak*) eragiten dio. Azkenik, (55) adibidean, *ezta* ezeztapen-markak sintagma osoari eragiten dio.

Azkenik, aurretik aipaturiko ezeztapen-markek sentimendu-balentzian eraginik ez duten kasuak ere badaude. Kasu hori sentimendu-balentzia indartzea baino ohikoagoa da, corpusean instantzien % 8,08an, 29 instantzietan, gertatu baita. Ezeztapen-markak berak ez du sentimendu-balentzian eraginik horrelako kasuetan: i) ezeztapen-marka juntagailuarekin agertzen denean, ii) ezeztapen-marka kontrastezko ezeztapenaren parte denean, eta iii) ezeztapen-marka egitura lexikalizatu batean agertzen denean (hots, egitura horiek beren berezko esanahia dutenean eta horietako batzuk hiztegi-tako sarrerak direnean).

- (56) Cate Le Bon bai edo ez, hemen ez dago erdibidekorik. (MUS33)
- (57) Ikuspuntu politikotik₋₁ ez ezik, ekonomikotik₊₃ ere Greziak esperantza ekarri du Europako hegoaldeko beste herrietara, tartean Euskal Herrira. (POL08)
- (58) Sei puntu baino ez dituela, hamaseigarren postuan da Realak sailkapenean. (KIR27)

(56) adibidean, idazleak aurkako bi ideia adierazten ditu (bai eta ez). Horrenbestez, ez da balentzia-aldaketarik sortzen. (57) adibidean, kontrastezko

ezeztapen bat dago eta informazioa gehitzeko funtzioa betetzen du (Silvennoinen, 2017). Hots, ezeztapen-marka emendiozko juntadura batean agertzen. Horrelako egituretan, *ez ezik* ezeztapen-markak bere aurretik dagoen elementua (*Ikuspuntu politikotik* adibidearen kasuan) ezeztatzen du. Baina, sentimendu-balentziaren ikuspegitik, ez da balentziarik ezeztatu eta ondorioz, ahuldu. Esaldian, ezeztapen-marka daraman lehen zatiak bere atzetik informazioa emendatzen edo gehitzen du. Azkenik, (58) egitura lexikalizatuaren adibidea da.

5.2.2. Trukaketa- eta desplazamendu-ezeztapena

Sentimenduen analisisian eta zehazki, lexikoian eta hizkuntza-ezagutzan oinarritzen den ikuspegian, bi hurbilpen daude ezeztapena lantzeko. Bi hurbilpen horiek trukaketa-ezeztapena (*switch negation*, ingelesez) (Sauri, 2008) eta desplazamendu-ezeztapena (*shift negation*, ingelesez) dira. Azter ditzagun bi hurbilpenak beheko bi adibideaetan³.

(59) Edaritegia [ez da ona₊₃]⁻³, baina bertako musika ona₊₃ da.

(60) Edaritegia [ez₋₄da ona₊₃]⁻¹, baina bertako musika ona₊₃ da.

Lehen adibideak (59) trukaketa-ezeztapenari dagokio eta bigarrenak (60), berriaz, desplazamendu-ezeztapenari. Lehen adibidean, *ez* ezeztapen-markak ez du balentziarik eta ezeztapen-markaren eragina da bere irismeneko parte den *ona* hitzaren sentimendu-balentzia (+3) alderantzikatzea. Beraz, trukaketa-ezeztapenean, irismeneko sentimendu-balentziari zeinua aldatzen zaio, positibotik negatibora edo alderantziz. Bigarren adibidean, desplazamendu-ezeztapenaren adibidea ageri da. Kasu horretan, *ez* ezeztapen-markak balentzia jakin bat du (-4) eta ezeztapen-markaren eragina da irismeneko hitzen sentimendu-balentziaren baturari (adibidean, -4) sentimendu-balentzia gehitzea. Irismeneko sentimendu-balentzia negatiboa bada, ezeztapen-markaren balentzia +4 izango dela.

Bi hurbilpen horien artean, guk desplazamendu-ezeztapena erabili dugu. Hurbilpen hori sintaxi mailako ezeztapenean eta ezeztapenarekin loturiko

³(59) eta (60) adibideak (Taboada *et al.*, 2011) lanetik hartu eta itzuli dira.

hizkiekin erabili dugu. Balentzia-aldatzaile horri ± 4 balentzia esleitu diogu eta, ondoren, balentzia duten hitzetan zer aldaketa eragiten duten aztertu dugu. Taboada *et al.*ek (2011) aipatzen duenez, azken hurbilpen hori hobeto dabil eta ez dago kontraesanik trukaketa-ezeztapenean bezala. Azter dezagun trukaketa-ezeztapenak duen akatsa.

(61) $Bikaina_{+5} \rightarrow Ez\ bikaina_{-5}. Anker_{-5}.$

(62) $Ez\ bikaina_{-5}. Ez\ ona_{-3}.$

(61) adibidean, trukaketa-ezeztapena erabiliz, *bikaina* adjektiboa +5 sentimendu-balentzia izatetik, ezeztapen-markaren eraginez, -5 sentimendu-balentzia izatea igarotzen da. Baina, hemen lehen kontraesana sortzen da; izan ere, intuitiboki *ez bikaina* eta *anker* elkarrengandik urrun daude, baina hurbilpen horren arabera, biek sentimendu-balentzia bera dute.

Bigarren kontraesana (62) adibidean beha daiteke. Bertan *bikaina* (+5) eta *ona* (+3) ezeztatuta daude. Ezeztapenaren eraginez, *ez bikaina* egiturak -5 sentimendu-balentzia du eta *ez onak* aldiz -3. Baina biak alderatzen baditugu, *ez bikaina ez ona* baino positiboagoa dela ohartuko gara. Hala ere, trukaketa-ezeztapena jarraituz, *ez bikaina* egiturak sentimendu-balentzia negatiboagoa du *ez onak* baino. Horrengatik guztiarengatik, mugimendu ezeztapena hurbilpena hautatu dugu eta ezeztapen-markek ± 4 balentzia izango dute.

5.2.3. Ezeztapen-markak eta beren irismena identifikatzeko erregelak

Metodologian, 3.2.2 atalean, ezeztapen-markek aditzen eta sintagmen sentimendu-balentzia nola eragiten duten identifikatu ondoren; ezeztapen-markak eta beren inguruneko hitzen gramatika-kategoriekin 5.6 taulako erregelak osatu ditugu. Erregela horiek *Murritzapen Gramatikan*, MGn, (Karlsson *et al.*, 1995) erregelak sortzeko baliatu dira.

Murritzapen Gramatika (Karlsson *et al.*, 1995) testuingurua iruditu zaigu egokiena ebaluazioa egiteko, gramatikaren ezaugarriak ondo uztartzen baitira

Zenb.	Ezeztapen-marka	Erregelen egitura
1	ezin	PM <i>ezin</i> + [adjektiboa/adberbioa] (+ atzizki konp.) PM
2	ez	PM [(IS +) aditza +] <i>ez</i> PM PM [(IS +)] <i>ez</i> [+ ad. lag. (+ IS) + aditza (+ IS)] PM PM [IS +] <i>ez</i> PM
3	ez	PM <i>ez</i> [+ IS +] <i>ez</i> [+ IS] (...) PM
4	gabe	PM [IS/AS/sintagma +] <i>gabe</i> PM
5	ezin	PM [(IS) + aditza + <i>ezin</i>] PM PM [<i>ezin</i> (+ ad. lag.) (+ IS) + aditza (+ IS)] PM
6	izan ezik	PM [IS/sintagma] + <i>izan ezik</i>
7	salbu	PM [IS] + <i>salbu</i> PM
8	ezta	PM <i>ezta</i> + [IS/sintagma] PM
9	ez	PM [aditza/bai] <i>edo/ala/edota ez</i> PM
10	ez	PM [IS] <i>ez ezik</i> PM
11	ez, gabe, ezin, ezean	Egitura lexikalizatuak

5.6 taula: Sentimendu-balentzian eragin ezberdinak dituzten ezeztapenak identifikatzeko proposatu diren erregeletako batzuk.

ebaluaratu nahi ditugun alderdiekin. Hiru ezaugarri nagusi eta abantaila ditu gramatika horrek:

- *Murriztapen Gramatika* (Karlsson *et al.*, 1995) hizkuntzarekiko independentea da. *Murriztapen Gramatika* analisi morfologikoan oinarritzen da, edozein testu analizatzeko helburuarekin. Gainera, egoera finituko mekanismoetan oinarritzen da, eta hori hizkuntzalaritzaren ikuspegitik gramatikaren formalismoa oso intuitiboa eta erabilerraza da.
- Desanbiguazio-erregelak. Gramatikako erregelek analisi morfologikoaren eta funtzio sintaktikoen desanbiguazioa gauzatzen dute. Erregelek testuinguru jakin batean zuzenak edo egokiak ez diren interpretazioak kentzen dituzte.
- Islapen-erregelak. Erregela horiek etiketatze morfologikoan eta sintaktikoan egon daitezkeen hutsuneak osatzen dituzte, testuingurua kontuan hartuz. Guk ezeztapen-markaren testuingurua, hau da, irismena identifikatu nahi dugu eta, horregatik, islapen-erregelak egokiak dira guretzat.

5.2.3.1. Murriztapen Gramatikako erregelak sortzeko eta ebaluatzeko prozedura

Erregelaren egituraketa

Jarraian urratsez urrats, 5.6 taulako erregelak ingurunera nola egokitu ditugun esplikatu dugu. *Murriztapen Gramatika* (Karlsson *et al.*, 1995) tes-tuinguruan sartu behar ditugun elementuak hauek dira⁴:

- Puntuazio-markak, ezeztapen-markak eta egitura lexikalizatueta-ko hitzak elementuen zerrendetan (LIST) sartzea erabaki dugu. Ia erregela denetan agertuko diren hitzak dira eta hitz jakinak dira, hots, beti forma bera dute.

(63) LIST EZ = "ez";

(64) LIST BESTERIK = "beste";

(63) eta (64) adibideetan, berriz, *ez* ezeztapen-marka eta egitura lexikalizatueta-ko *bestestik* hitza, hurrenez hurren, elementuen zerrendetako elementu bakarrak dira, modu horretan, hitz horiek ezeztapen-marken erregela ezberdinekin konbinatzeko aukera edukitzeko.

- Irismeneko hitzen gramatika-kategoriak eta haien posizioa islapen-erregelatan sartu ditugu. Irismenean, aurreko kasuetan ez bezala, ez dakigu zein hitz ager daitezkeen bertan. Horregatik, ezin ditugu irismeneko hitzekin elementuen zerrendak osatu. Dakigun bakarra irismenean ager daitezkeen hitzen gramatika-kategoria eta ezeztapen-marketatik haien distantzia dira, 5.6 taularen bidez lortu ditugunak. Horrenbestez, informazio hori islapen erregelatan adierazi dugu:

(65) MAP (!salbuBUK) TARGET (ADB) IF (0C SALBU);
MAP (!salbuHAS1) TARGET (IZE) IF (1C SALBU);

(65) adibidean, bestalde, bi erregela ageri dira *larunbata salbu* bezalako egiturak identifikatzeko. *!salbuBUK* etiketa duen erregelaren, 0 posizioan SALBU elementuen zerrendak egon behar duela adierazten du

⁴Elementu horiek ikusgai daude A eranskinean.

eta, gainera, elementu zerrendakoak adberbioa izan behar du. Bestalde, *IsalbuHAS1* etiketan duen erregelari, *salbu* ezeztapen-markaren aurretik izen batek egon behar duela adierazten da. Kasu horretan, hitza edozein izan daiteke, baina badakigu izenaren gramatika-kategoriakoak izan behar duela, 5.6 taularen bidez. Aipaturiko etiketa horiek izango dira sortu ditugun erregelak corpusetik pasatzean utziko duten aztarna.

Erregelak ebaluatzeko prozedura

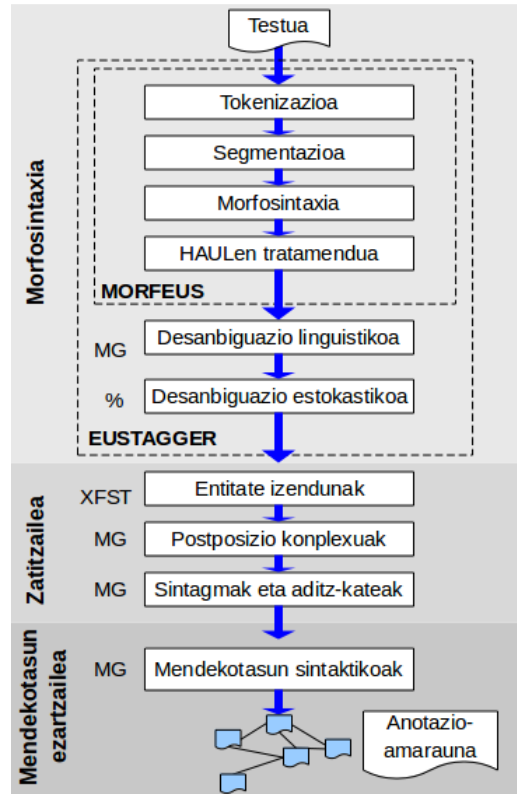
Erregelen ebaluazioa etiketatzaile batek gauzatu du, baina, alde aurretik, etiketatzaile horren eta beste etiketatzaile baten adostasun maila neurtu dugu, ebaluazioaren kalitatea ebaluatu ahal izateko. Bi etiketatzailek Euskarazko Iritzi Corpusaren (Alkorta *et al.*, 2016) % 10a (2.706 hitz) etiketatu dute. Bi etiketatzaileen arteko adostasuna zein hitz etiketatzeari dagokionez, kappa 0,91koa da. Bi etiketatzaileak etiketatutako beharreko hitzak nola etiketaturik (ETIK_ONDO, ETIK_FALTA eta ETIK_GAIZKI) dagokionez, kappa 0,69koa da, zeina Landis eta Kochen (1977) arabera *sendoa* den.

Bi etiketatzaileen arteko adostasuna neurtu ostean, Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) 144 iritzi-testu erabili ditugu *Murriztapen Gramatikan* (Karlsson *et al.*, 1995) sortutako erregelak ebaluatzeko. 144 iritzi-testuko corpusari Ixa taldearen analisi-katea (Aduriz *et al.*, 2004) pasa diogu. Analisi-katearen arkitektura 5.2 irudian ikus daiteke.

Ixarko analisi-katearen oinarrian EDBL-LBDBL baliabide lexikal konputazionala (Aldezabal *et al.*, 2006) dago eta analisi-kateak hiru atal ditu: morfosintaxia, zatitzailea eta mendekotasun ezartzailea. Corpusari analisi-katea pasa ondoren, emaitza hitz bakoitzeko analisi bakarra izan da 5.3 irudian agertzen den moduan.

Bertan, analisi-katearen arabera, *azaroaren* hitza izen (IZE) arrunta (ARRU) da eta bere kasua genitiboa (GEN>) da.

Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) 144 iritzi-testuei analisi-katea pasa ostean, *Murriztapen Gramatikan* (Karlsson *et al.*, 1995) sortutako erregelak aplikatu ditugu corpusean eta erregela horiek corpuseko analisisetan beren etiketa utzi dute. 3.2.2 atalean aipatu dugun moduan, etiketa erregelak ebaluatzeko erabili ditugu.



5.2 irudia: Ixa taldearen analisi-katea (Ornoz, 2009).

```
"<azaroaren>"
<Correct!> "azaro" IZE ARR GEN NUMS MUGM ZERO w67,L-A-IZE-ARR-26,
lsfi6 @IZLG>
```

5.3 irudia: Ixa taldeko analisi-katearen emaitza.

Murritzapen Gramatika (Karlsson *et al.*, 1995) esleitutako etiketak zuzenak edo okerrak diren, edo etiketaren bat falta den ebaluatzeko hiru kasuistika bereiztu ditugu⁵:

- ETIK_ONDO. Sorturiko erregelak dagokion hitzari etiketa jarri badio ontzat joko dugu. Kasu horretan, beraz, erregelak ondo identifikatu du

⁵Erregelak ebaluatzeko etiketak corpusaren zati batean aplikatuta 3.9 irudian daude ikusgai.

ezeztapenarekin loturiko elementuren bat (ezeztapen-marka, irismena edo egitura lexikalizatua) corpusean.

- **ETIK_FALTA.** Sorturiko erregelak dagokion hitzari etiketa ez badio jartzen okertzat joko dugu. Beste kasu batzuetan, gerta liteke erregelak ezeztapenarekin loturiko elementuren bat ez identifikatzea corpusean eta kasu horretan erabiliko dugu etiketa hori.
- **ETIK_GAIZKI.** Sorturiko erregelek ez dagokion hitzari etiketa jartzen badio okertzat joko dugu. Azken kasuan, gerta liteke erregelek ezeztapenarekin loturirik ez duen hitz bat ezeztapenarekin lotura duen elementutzat jotzea eta, kasu horretan ere, kasuistika hori okertzat joko dugu.

5.2.3.2. Erregelen ebaluazioaren emaitzak

Lehenik eta behin, emaitzak ezeztapen-markek sentimendua balentzian eragiten dutena (indartu, ahuldu edo eraginik ez) kontuan hartuta emango ditugu. Ondoren, emaitzak ezeztapenari lotutako elementuetan (ezeztapen-markak, irismena eta egitura lexikalizatuak) arabera emango ditugu.

5.7 taulako datuen arabera, erregelek ezeztapen-markekin lotutako osagaien hitzak ondo identifikatzen ditu, F1 puntuazioa 0,86 baita. Zehatzago esanda, erregelek ezeztapenarekin lotutako hitzak identifikatzerakoan 0,91ko doitasuna eta 0,80ko estaldura izan dute.

Eragina	Doitasuna	Estaldura	F1	Hitzak (instantziak) ✓/×/✚
Indartu	1,00	1,00	1,00	4 (2): 4/0/0
Ahuldu	0,93	0,80	0,86	1.050 (192): 784/63/203
Eraginik ez	0,97	1,00	0,98	36 (19): 35/1/0
Guztira	0,93	0,80	0,86	1.090: 823/64/203

5.7 taula: Emaitzak ezeztapen-markek sentimendu-balentzian eragiten dutenaren ikuspegitik.

Ezberdintasun batzuk daude F1 puntuazioan, ezeztapen-markek balentzian eragiten dutenaren arabera. Esaterako, ezeztapen-markak indartzen duenean edota eraginik ez duenean, F1 puntuazioa altua da. Baina, eurei loturiko

hitz eta instantzia⁶ kopurua baxua da. Ezeztapenak balentzia ahultzearen kasuan, aldiz, F1 puntuazioa 0,86 da, baxuagoa. Indartzearen kasuan, ezeztapenari lotutako hitz guztiak (4) ondo identifikatu dira. Ahultzean, 784 ondo identifikatu dira, 63 gaizki eta 203 ez dira identifikatu. Eraginik ez dagoen kasuetan, 35 hitz ondo identifikatu dira eta bakarra gaizki. Ezeztapena ahultze eraginaren kasuan, F1 puntuazioa 0,86 da. Emaitzak ezeztapenari loturiko elementu-motan oinarrituz ageri da 5.8 taulan.

	Doitasuna	Estaldura	F1	Hitzak (instantziak): ✓/×/÷
Ezeztapen-markak	1,00	0,96	0,98	195 (195): 188/0/7
Egitura lexikalizatuak	0,96	1,00	0,98	28 (16): 27/1/0
Irismena	0,91	0,75	0,82	867 (195): 601/63/196
Guztira	0,93	0,80	0,86	1.090: 823/64/203

5.8 taula: Emaitzak ezeztapenari loturiko elementu-motaren ikuspegitik.

Bertan beha daitekeenez, ezeztapen-markak eta egitura lexikalizatuak oso ondo identifikatu dira, F1 puntuazioa 0,98 baita bietan. Zazpi ezeztapen-marka ez dira identifikatu eta corpuseko hitz bat egitura lexikalizatutzat hartu da, nahiz eta ez den horrela. Irismenean, ordea, F1 puntuazioa baxuagoa da, 0,82koa, hain zuzen ere. Irismenaren parte diren 196 hitz ez dira identifikatu eta irismenaren parte ez diren 64 hitz irismeneko zatitzat hartu dira. Horretaz gain, 867 hitzetatik 196 ez dira irismeneko parte bezala jo, nahiz eta hala diren.

Errore-analisiak erregelen osaketari buruzko informazioa eman digu. Irismenari dagokionez, erregelek ez dituzte 196 hitz identifikatu (identifikatu gabeko zazpi ezeztapen-markak kontuan hartu gabe). Hitz horiek zer dela eta identifikatu gabe gelditu diren aztertu ondoren, kasu gehienetan, puntuazio-markaren murriztapenarekin lotura dutela ohartu gara. Hau da, puntuazio-marka murriztapena irismena esaldi barrura mugatzeko sortua izan da, baina

⁶5.7 taulan, eskuineko zutabean, hitzei eta instantziei loturiko datuak jarri ditugu. Lehenik eta behin, hitz-kopurua agertzen da, ondoren, parentesi artean, instantzia-kopurua eta ondoren, ezeztapenari loturiko hitzak ondo, gaizki edo ez diren asmatu. Ikusten den moduan, hitz eta instantzia-kopuruak ezberdinak dira. Ezeztapen-marken kasuan, hitz bat instantzia bat da. Irismenaren kasuan, instantziako hitz-kopurua ezberdina da. Guztira 192 irismen daude corpusean eta horiek 1.050 hitz dituzte. Azkenik, egitura lexikalizatuetan, hitz batetik hiru hitzerainokoa izan daiteke instantzia bat, betiere egitura lexikalizatu horren ezaugarrien arabera.

horrek zeharkako kalte batzuk sortzen ditu. Esaldiren batean zerrendaketa bat dagoenean, eta zerrendaketa hortako elementuak koma bidez bereizita daudenean, puntuazio-markaren murriztapena ez da ondo ibiltzen, zerrendaketako elementu bakarra identifikatzen baitu, esaldia hor bukatzen dela interpretatzen duelako.

5.3 Diskurtso maila

5.3.1 azpiatalean, diskurtso mailan aztertu ditugun balentzia-aldatzaileak aurkeztu ditugu. 5.3.2 atalean, Egitura Erretorikoaren Teoria (RST) (Mann eta Thompson, 1988) deskribatuko dugu.

5.3.1. Balentzia-aldatzaileak: nukleartasuna eta unitate zentrala

Balentzia-aldatzaileen azterketa diskurtso mailan hiru ikuspegitatik egin dugu. Alde batetik, nukleartasunak balentzia-aldaketan eraginik duen aztertu dugu (5.3.1.1 atala). Beste aldetik, unitate zentralak balentzia-aldaketan eraginik ere baduen aztertu dugu (5.3.1.2 atala). Azkenik, unitate zentralak Azkenik, erlaziozko diskurtso-egiturak unitate zentraletik duen distantzia baikoitzean zenbateko maiztasunez agertzen diren aztertu dugu (5.3.1.3 atala).

5.3.1.1. Orientazio semantikoa eta nukleartasuna

Balentzia-aldatzaileen eta nuklearitatearen arteko lotura aztertzeke diskurtso-erlazioak aztertu ditugu. Zehazki, zer diskurtso-unitatek (nukleoak edo sateliteak edota lehen testu-zatia edo azken-testu zatia) duen erlazio osoarekin orientazio semantikoaren adostasun gehien neurtu dugu. 5.9 taulan, orientazio semantikoaren adostasunari buruzko emaitzak azter daitezke, nuklearitate eta osagaien lekua aintzat hartuz⁷. Emaitza horietan, adostasunik handiena +1 balioa da eta adostasunik baxuena 0 balioa.

5.9 taulako emaitzen arabera, kasu gehienetan, nukleo-unitateak erlaziozko diskurtso-egitura osoaren orientazio semantikoarekin adostasun handiagoa erakusten du satelite-unitateak baino. Orientazio semantikoaren adostasunean nukleo-unitateak eta satelite-unitateak duten ezberdintasuna ez da oso handia. 0,73 da nukleo-unitatearen adostasuna eta 0,64, berriz, satelitearena. Hala ere, joera hori ez da gertatzen erlaziozko diskurtso-egitura mota guztietan. Ebaluazioa taldearen⁸ kasuan, satelite-unitateak erlaziozko

⁷Letra xehez dauden erlaziozko diskurtso-egiturak multzoak dira. Letra larriz daudenak erlaziozko diskurtso-egitura bakarra dira.

⁸Ebaluazioa taldea EBALUAZIOA eta INTERPRETAZIOA erlaziozko diskurtso-

Adostasuna Xrekin	Instantziak	Nukleoa	Satelitea
Kausa taldea	71	0,83	0,66
Aurkaritza taldea	70	0,70	0,33
BALDINTZA	18	0,61	0,56
Ahalbideratzea taldea	6	0,60	0,5
Ebaluazioa taldea	140	0,69	0,82
Ebidentzia taldea	53	0,83	0,64
KONTRASTE*	26	0,73	0,31
Guztira	384	0,73	0,65

5.9 taula: Orientazio semantikoaren adostasuna erlaziozko diskurtso-egituren eta nukleoren/satelitearen artean.

diskurtso-egitura osoaren orientazio semantikoarekin adostasun handiagoa erakusten du satelite-unitateak baino. Kasu horretan, satelite-unitatearen adostasuna 0,82koa da nukleo-unitatearena 0,69 den bitartean. Bestalde, Aurkaritza taldeak⁹ du nukleo-unitatearen eta satelite-unitatearen arteko ezberdintasunik handiena. Nukleo-unitateak 0,70ko adostasuna du eta satelite-unitateak, aldiz, 0,33. Beraz, puntuazioan 0,44ko ezberdintasuna dago. Bukatzeko, nukleo-unitateak Kausa taldean¹⁰ eta Ebidentzia taldean¹¹ erlaziozko diskurtso-egitura osoaren orientazio semantikoaren adostasunik handiena dute, puntuazioa 0,83koa baita bietan.

Erlaziozko diskurtso-egitura osoaren orientazio semantikoaren adostasunik handiena lehen eta azken unitateak duen aztertzen badugu, emaitzak argigarriagoak dira. 5.10 taulak erakusten duen moduan, erlaziozko diskurtso-egituraren azken unitateak erlazio osoaren orientazio semantikoarekin adostasun handiagoa du. Azken unitatearen kasuan, adostasunaren puntuazioa 0,78koa da eta lehen unitatearen kasuan, berriz, 0,58koa. Beren artean dagoen aldea esanguratsua da. Kasu horretan ere, erlaziozko diskurtso-egitura mota guztiek ez dute joera erakusten. Ahalbideratzea¹² taldean, erlazioa-

egiturek osatzen dute.

⁹Aurkaritza taldea KONTZETSIOA eta ANTITESIA erlaziozko diskurtso-egiturek osatzen dute.

¹⁰Kausa taldea KAUSA, ONDORIOA eta HELBURUA erlaziozko diskurtso-egiturek osatzen dute.

¹¹Ebidentzia taldea EBIDENTZIA eta JUSTIFIKAZIOA erlaziozko diskurtso-egiturek osatzen dute.

¹²Ahalbideratzea taldea AHALBIDERATU eta MOTIBAZIOA erlaziozko diskurtso-

Adostasuna Xrekin	Instantziak	Lehen unitatea	Azken unitatea
Kausa taldea	71	0,66	0,83
Aurkaritza taldea	70	0,30	0,73
BALDINTZA	18	0,28	0,89
Ahalbideratzea taldea	6	0,67	0,50
Ebaluazioa taldea	140	0,68	0,83
Ebidentzia taldea	53	0,79	0,68
KONTRASTE*	26	0,35	0,69
Guztira	384	0,58	0,78

5.10 taula: Orientazio semantikoaren adostasuna erlaziozko diskurtso-egituren eta lehen osagaiaren/azken osagaiare artean.

ren lehen unitateak azkenak baino orientazio semantikoaren adostasun handiagoa dute. Ahalbideratzearen kasuan, lehen unitatearen adostasun puntuazioa 0,67koa da eta azken unitatearena 0,50ekoa. Bestalde, Ebidentzia taldean puntuazioak 0,79 eta 0,68 dira, hurrenez hurren. Lehen unitatearen eta azken unitatearen artean, puntuazioaren ezberdintasunik handiena BALDINTZA erlazioan gertatu da. Orobat, lehen unitatearen puntuazioa 0,28koa da eta azken unitatearena 0,89. Hots, 0,71ko puntuazioaren tarte dago. Azkenik, BALDINTZA erlazioko, Ebaluazioa taldeko eta Kausa taldeko azken unitateek erlaziozko diskurtso-egitura osoaren orientazio semantikoarekin adostasun puntuaziorik handiena dute, puntuazioa 0,89koa baita BALDINTZAREN kasuan eta 0,83koa Ebaluazioarenean eta Kausarenean.

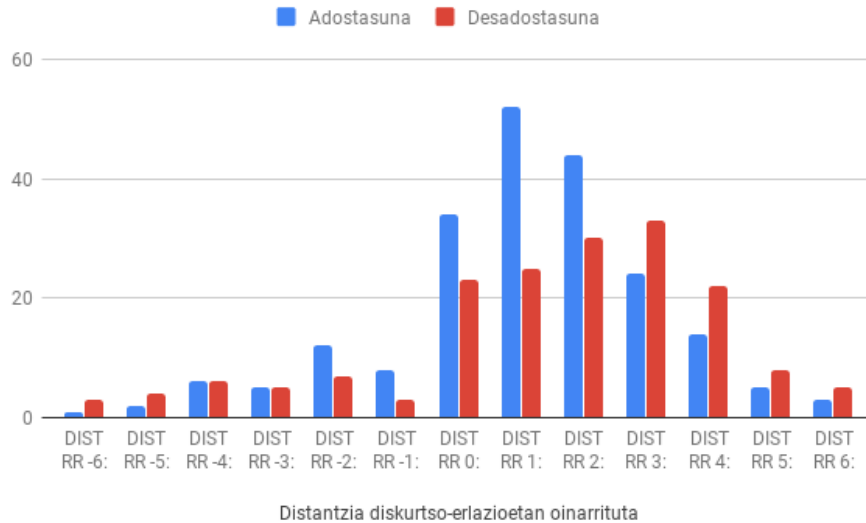
Laburbilduz, erlaziozko diskurtso-egituraren orientazio semantikoa baldintzatzen dute bai nukleartasunak, bai lehen edo azken unitatea izateak ere.

5.3.1.2. Orientazio semantikoa eta unitate zentrala

5.4 irudian, erlaziozko diskurtso-egitura guztiek beren testu osoarekin duten orientazio semantikoaren adostasun eta desadostasun instantziak agertzen dira. Unitate zentrala (DIST RR 0) da eta erlaziozko diskurtso-egiturak bere aurretik eta atzetik daude.

Erlaziozko diskurtso-egitura unitate zentraletik zenbat eta gertuago egon, erlaziozko diskurtso-egituren eta beren iritzi-testuen arteko orientazio semanti-

egiturek osatzen dute.



5.4 irudia: Testuek eta erlaziozko diskurtso-egiturek beren artean dituzten orientazio semantikoaren adostasun eta desadostasunak.

koaren adostasuna handiagoa da. -2 , -1 , 0 , $+1$ eta $+2$ distantzietan orientazio semantikoaren adostasunaren instantzia gehiago daude desadostasunak¹³ baino. Beste distantzietan, ordea, orientazio semantikoaren desadostasunaren instantziak gehiago dira edo bestela adostasun eta desadostasunen artean ez dago ezberdintasun nabarmenik.

Emaitzak erlaziozko diskurtso-egitura moten arabera aztertzen badira, zenbait ezberdintasun antzeman daitezke. Kausa taldearen kasuan, orientazioaren adostasunaren instantziak gehiago dira 0 eta $+1$ distantzietan bakarrik. Aurkaritza taldeak, aldiz, ez du erlaziozko diskurtso-egiturek distantzietan erakusten duten joera jarraitzen. Izan ere, Aurkaritza taldean, adostasun instantziak bi gunere bereizitan dira nagusi. Hau da, adostasuna gunere batean nagusi da (-2 distantzia); ondoren, desadostasun instantziak gehiago dira (0 distantzia) eta, azkenik, adostasun instantziak berriz nagusitzen dira ($+1$ eta $+2$ distantziak).

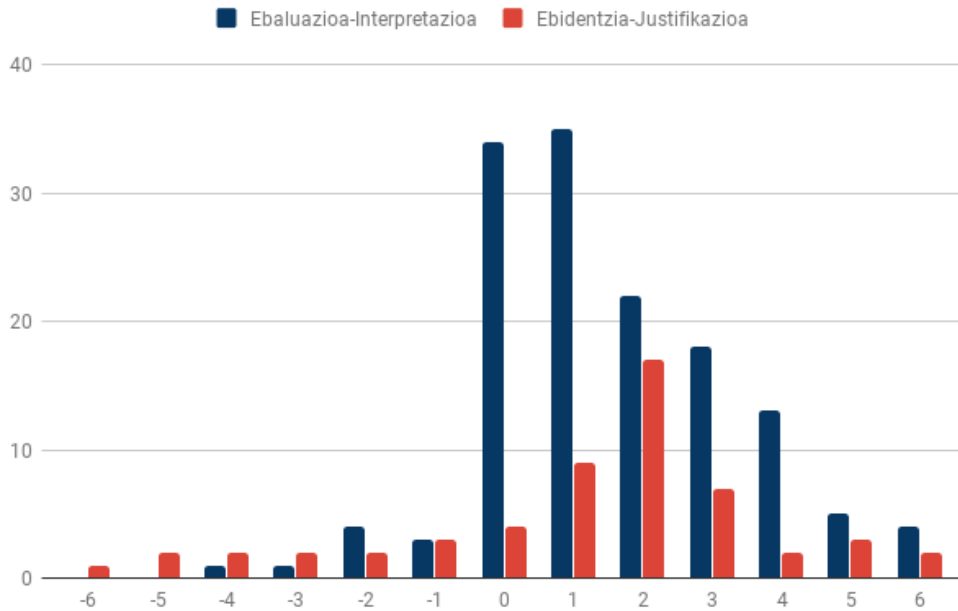
¹³Desadostasuna izendapenez, erlaziozko diskurtso-egiturak eta iritzi-testuak aurkako orientazio semantikoa duten instantziak adierazi nahi ditugu.

Laburbilduz, unitate zentralak erlaziozko diskurtso-egiturak eta iritzi-testuak duten orientazio semantikoaren adostasunean eragiten du, izan ere, erlaziozko diskurtso-egitura bat unitate zentraletik gertuago egon ahala, erlazioak duen orientazio semantikoak iritzi-testuak duenarekin gehiagotan bat egiten du. Gure ustez, unitate zentraletik gertuen dauden erlaziozko diskurtso-egituren gaiak iritzi-testuaren gai nagusitik hurbilago daudelako dago bien arteko orientazio semantikoaren adostasun handiagoa. Unitate zentraletik urrun dauden erlaziozko diskurtso-egituren, ordea, gaia iritzi-testuarengandik ezberdinak dira eta, ondorioz, orientazio semantikoan adostasuna ez da hain bestekoa.

5.3.1.3. Erlaziozko diskurtso-egituren agerpena distantzia bakoitzean

Iritzi-testuetan, erlaziozko diskurtso-egiturak distantzia bakoitzean zenbateko maiztasunez agertzen diren ere neurtu dugu. Ebaluazioa-Interpretazioa eta Ebidentzia-Justifikazioa erlaziozko diskurtso-egituraren taldeen emaitzak aurkeztuko ditugu. Distantzia unitate zentraletik neurtu dugu. Distantzia negatiboa da unitate zentralaren ezkerretara eta positiboa unitate zentralaren eskuinetara. 5.5 irudian, Ebaluazioa-Interpretazioa eta Ebidentzia-Justifikazioa erlaziozko diskurtso-egituraren taldeek instantzia gehien zer distantzietan duten beha daiteke.

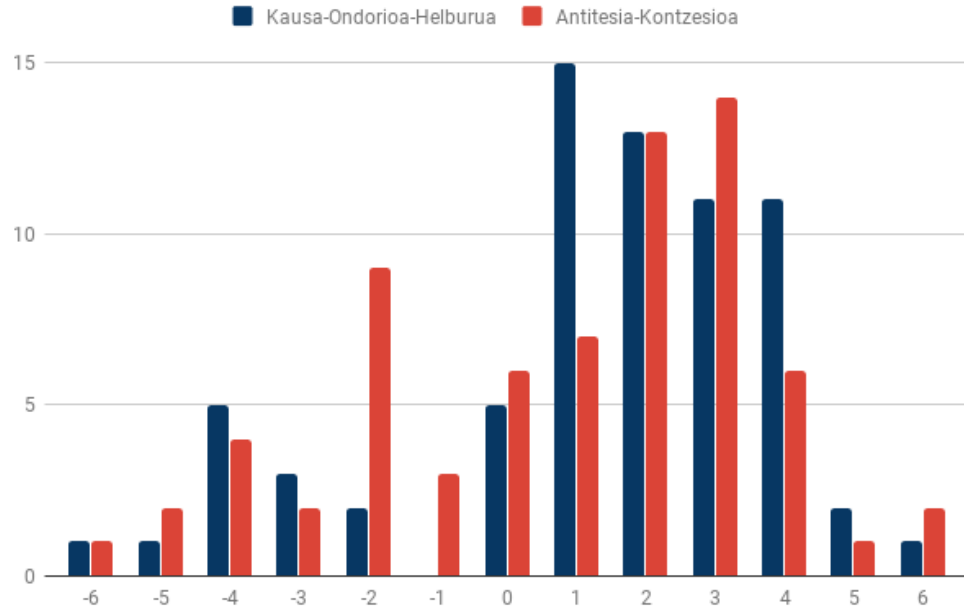
Ebaluazioa-Justifikazio erlaziozko diskurtso-egituraren taldearen instantzia gutxi daude unitate zentralaren aurretik, baina, 0 eta +1 distantzietan, bere instantzia kopurua asko igotzen da. Ondoren, berriz, instantzia kopurua jaitsi egiten da, unitate zentralaren aurreko egoerara itzuliz. KONTRASTEIA erlaziozko diskurtso-egitura multinuklearrak ere modu berean jokatzen du. Ebidentzia-Justifikazioa erlaziozko diskurtso-egituraren taldean, ordea, distantzia guztietan dago erlaziozko diskurtso-egituraren talde horren instantziaren bat eta +1 eta +3 distantzien artean, bere instantzien kopurua nabarmen igotzen da. Beraz, Ebaluazioa-Interpretazioa erlaziozko diskurtso-egituraren taldeak distantzietan agerpen irregularra du eta Ebidentzia-Justifikazioa erlaziozko diskurtso-egituraren taldeak, aldiz, agerpen erregularra. Bestalde, 5.6 irudiak Kausa-Ondorioa-Helburua eta



5.5 irudia: Ebaluazioa-Interpretazioa eta Ebidentzia-Justifikazioa erlaziozko diskurtso-egituren instantziak.

Antitesia-Kontzetsioa erlaziozko diskurtso-egituraren taldeen emaitzak erakusten ditu.

Bi erlaziozko diskurtso-egitura taldeek ezaugarri batzuk partekatzen dituzte, unitate zentraletik duten instantzia handieneko guneei dagokienez. Ebidentzia-Justifikazioa erlaziozko diskurtso-egituraren taldean bezala, erlazio horiek unitate zentralarekiko distantzia guztietan agertzen dira. Baina erlazio taldeek instantzien kuantitatean bi goreneko guneei dituzte. Bi multzo horietan, goreneko uneeetako bat unitate zentralaren aurretik dago eta bestea unitate zentralaren ondoren. Antitesia-Kontzetsioa erlazioaren kasuan, bi goreneko uneeak -2 eta $+3$ distantzietan daude. Instantzia maiztasunen bi goreneko uneeak -4 eta $+1$ distantzietan agertzen dira Kausa-Ondorioa-Helburua erlaziozko diskurtso-egituraren taldean.



5.6 irudia: Kausa-Ondorioa-Helburua eta Antitesia-Kontzesioa erlaziozko diskurtso-egitura taldeen instantziak.

2 Ikerketa hipotesia

1.2 ataleko ikerketa hipotesiari erantzunez, euskararen hizkuntza maila ezberdinetan testuinguruko balentzia-aldatzaileak daude.

Ezeztapena eta diskurtso-egiturako balentzia-aldatzaileak beste hizkuntzetan ere badaude, baina euskararen berezitasuna da morfologiazko balentzia-aldatzaileetan duen aberastasuna.

3 Ikerketa hipotesia

1.2 ataleko hipotesiari erantzunez, erlaziozko diskurtso-egituran nukleartasuna eta iritzi-testuetan unitate zentrala daude erlaziozko diskurtso-egituraren edota bertako osagaien orientazio semantikoa zeinu jakin batekoa izatea eragiten dutenak.

Nuklearitateak eta erlaziozko diskurtso-egituretakoko azken testuzatiak orientazio semantikoaren adostasun handiagoa dute. Beraz, elementu horiek diskurtso-unitateek erlaziozko diskurtso-egituraren orientazio semantikoa duten adostasuna indartu egiten dute.

Unitate zentralak testuko erlaziozko diskurtso-egiturek iritzi-testuaren orientazio semantikoarekin adostasun handiagoa izatea eragiten du. Erlaziozko diskurtso-egitura bat unitate zentraletik gertuago egon ahala, bere eta iritzi-testuaren orientazio semantikoaren adostasuna handiagoa da, hots, adostasuna indartu egiten du.

5.3.1.4. Unitate zentralaren azterketa

Unitate zentralaren azterketan, i) ezaugarri sintaktikoak aintzat hartuta domeinu bakoitzeko unitate zentralaren karakterizazioa eta ii) sentimendubalentziadun hitzen banaketa domeinu ezberdinetako unitate zentraletan aurkeztuko ditugu.

Unitate zentralen ezaugarri sintaktikoak

5.11 taulan, domeinu bakoitzeko unitate zentralen maiztasun handienez agertzen diren gramatika-kategoriak ikus daitezke¹⁴.

Domeinuka iritzi-testuen unitate zentralak dituzten ezaugarriak identifikatu ostean, domeinuen artean zenbait berdintasun antzeman ditugu. Eguraldiaren domeinuko unitate zentralak beste domeinukoetatik ezberdinak dira hiru alderditan. Batetik, eguraldiaren domeinuak, denbora aipatzea ohikoa da (esaterako: *gaur*, *bihar* edo *asteburua*). Bestetik, beste domeinuetan ez bezala, ez da entitaterik agertzen bertan. Azkenik, aditza ez agertzea ere

¹⁴Gramatika-kategoriez gain, entitateak, denbora eta determinatzailea ere kontuan hartu ditugu unitate zentralako hitzak maiztasunaren arabera aztertzean, mota horretako hitzak asko agertu direlako.

	Ize.	Adj.	Adi.	Adb.	Det.	Entitateak
Eguraldia	X	X		X		
Kirola			X		X	X
Musika	X		X			X
Politika		X				X
Filmak	X	X	X			X
Liburuak	X	X	X		X	X

5.11 taula: Unitate zentraletan bereizgarri diren ezaugarriak domeinuen eta gramatika-kategorien arabera.

ezaugarri bereizle bat da. Kirolaren domeinuari dagokionez, bertan determinatzaile ugari (partidetako emaitzak) agertzen dira. Aditzean zer entitatek irabazi edo galdu egin duen zehazten da.

Musika, film eta liburuetakoko domeinuek zenbait antzekotasun partekatzen dituzte. Izenak, adjektiboak, aditzak eta entitateak azaltzen dira unitate zentraletan. Azkenik, politikan, entitateak eta adjektiboak beste gramatika-kategoriak baino dezentez gehiago agertzen dira eta beste domeinuetatik, kasu horretan, ezberdindu egiten da.

Datu horiek aintzat hartuz, iritzi-testuen domeinuak unitate zentraletan gramatika-kategoria mota bat bestea baino gehiagotan agertzea eragiten du. Eguraldia eta kirolak ezaugarri bereizleak dituzte unitate zentralean eta beste domeinuek tarteko ezaugarriak dituzte. Unitate zentralen azterketa hori baliagarria izan daiteke etorkizunean iritzi-testuetako unitate zentralak antzemateko tresna edo baliabideak sartzeko.

Sentimendu-balentziadun hitzen banaketa unitate zentraletan

Corpuseko 192 unitate zentraletako hitzei sentimendu-balentzia euskarazko sentimenduen sailkatzailea erabiliz esleitu zaie. 5.12 taulak esleipen horren emaitzak erakusten ditu.

Corpuseko unitate zentraletan, 2.081 hitzetatik 258 hitzek (% 12,40ak) dute sentimendu-balentzia. Gramatika-kategorien ikuspegitik, 99 adjektiboek dute sentimendu-balentzia, 77na izen eta aditzek eta azkenik, bost adverbik. Datuak ehunekoetan emanda, adjektiboek sentimendu-balentziadun hitzen % 38 hartzen dute; izen eta aditzek % 30 eta azkenik, adverbioek % 2.

		%
Hitzak guztira	2.081	
Balentziadun hitzak guztira	258	12,40
<i>Adjektiboak</i>	<i>99</i>	<i>38</i>
<i>Aditzak</i>	<i>77</i>	<i>30</i>
<i>Izenak</i>	<i>77</i>	<i>30</i>
<i>Adberbioak</i>	<i>5</i>	<i>2</i>

5.12 taula: Euskarazko sentimenduen sailkatzaileak sentimendu-balentzia esleitutako unitate zentralako hitzak gramatika-kategoriaren arabera sailkatuta.

Beraz, adjektiboetan sentimendu-balentziadun hitz gehiago agertzen diren arren, banaketa nahiko orekatua da. Domeinutik domeinura dauden ezberdintasunak ikus daitezke 5.13 taulan.

	Eguraldia		Kirola		Literatura		Musika		Filmak		Politika		Guztira	
Izena	7	0,16	11	0,33	17	0,36	10	0,26	12	0,33	20	0,33	77	0,30
Aditza	12	0,27	16	0,48	12	0,25	13	0,34	10	0,28	14	0,23	77	0,30
Adjektiboa	22	0,50	5	0,15	18	0,38	15	0,39	14	0,39	25	0,42	99	0,38
Adberbioa	3	0,07	1	0,03	-	-	-	-	-	-	1	0,02	5	0,02
Guztira	44		33		47		38		36		60		258	

5.13 taula: Sentimendu-balentziadun hitzen agerpena.

Sentimendu-balentziadun hitzen kopuruan, politikaren domeinuan 60 sentimendu-balentziadun hitz agertzen dira eta kirolean, sentimendu-balentziadun hitzak 33 dira. Eguraldiaren domeinuan, adjektiboaren gramatika-kategoriak nabarmenki balentziadun hitzen agerpen gehiago ditu (sentimendu-balentzia duten hitzen erdiak adjektiboak dira). Kirolaren domeinuan, aditza da sentimendu-balentziadun hitz gehien dituen gramatika-kategoria (aditzak % 48 dira). Literaturan, musikan, filmeetan eta politikan adjektiboak dira sentimendu-balentziadun hitz gehien dituztenak. Adjektiboak balentziadun hitzen % 38-42 hartzen dute, baina beste gramatika-kategoriak datu horietatik gertu daude. Laburbilduz, balentziadun hitzak dituen gramatika-kategoria ohikoena adjektiboena da, indar handiagoz eguraldian eta oso indar gutxiz kirolean non bertan aditza nagusitzen den modu nabarmen batean.

Atal honetako emaitzak eta 5.3.1.4 atalean lortutakoak, sentimenduen analisira begira unitate zentralak identifikatzeko, beren sentimendu-balentzia

eta orientazioa semantikoa kalkulatzeko eta domeinu bakoitzeko unitate zentralen zer gramatika-kategoriari eman behar zaion garrantzia antzemateko balio dezake. Gerora, bildutako informazio hori baliagarria izan daiteke euskarazko sentimenduen sailkatzailean unitate zentrala lantzeko modulu bat sortzeko.

4 Ikerketa hipotesia

1.2 ataleko hipotesiari erantzunez, domeinutik domeinura unitate zentralak ezaugarri bereziak dituela egiaztatu dugu.

Kirolaren domeinuan aditzak garrantzi handia du eta eguraldiaren domeinuan, aldiz, adjektiboak. Entitateak ere garrantzitsuak dira hainbat domeinutan.

Sentimendu-balentzia duten gramatika-kategorien kasuan, kirolaren domeinuan, aditzak ohikoenak dira eta beste gramatika-kategorietan adjektiboak.

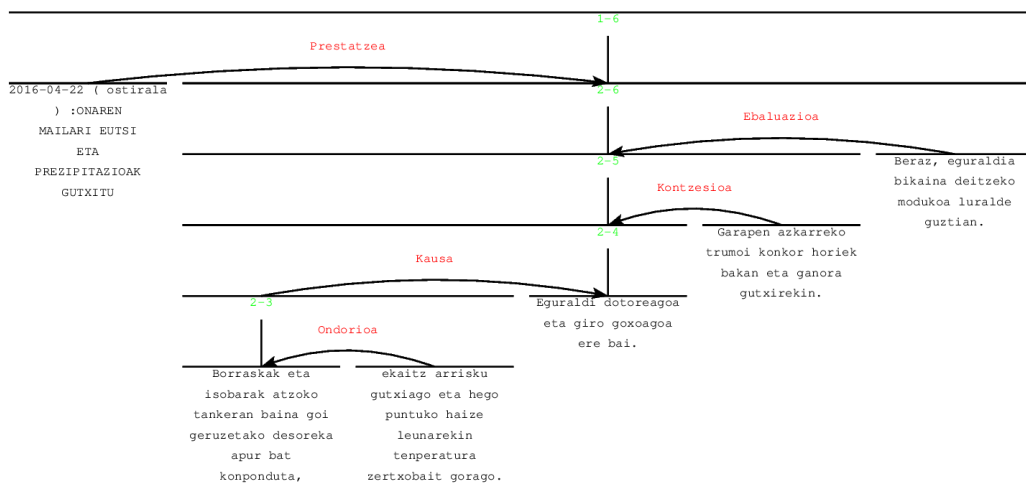
5.3.2. Egitura Erretorikoaren Teoria (RST)

Lan honetan, diskurtso-egitura aztertzeko Egitura Erretorikaren Teoria (*Rhetorical Structure Theory* edo RST) (Mann eta Thompson, 1988) hurbilpena hautatu dugu. Teoria honek testuaren egitura edo koherentzia deskribatzen du. Hizkuntzalaritza Konputazionalaren diziplinan eta erlaziozko diskurtso-egituran koherentzia aztertzeko gehien erabiltzen den teoria da eta euskaraz dagoen bakarrenetakoa da.

Hizkuntzalaritza Konputazionalan diskurtso-egitura lantzean bi fenomeno landu ohi dira: i) erreferentziazko fenomenoak: koerreferentziaren ebazpena lantzen da eta ii) erlaziozko fenomenoak: koherentzia erlazioen esleipena lantzen da. Egitura Erretorikoaren Teorian koherentzia erlazioak dira aztergai, baita gure lanean ere.

Mann eta Thompsonen (1988) RST konputagailu bidez testu-sorkuntza (*text generation*) egiteko asmoz sortu zuten. Testu-sorkuntza egin ahal izateko, testuaren ezaugarriak zehaztu behar dira eta testuen ezaugarririk garran-

tzitsuena koherentzia da. RSTk testuaren koherentzia deskribatzen duenez, konputagailua gauza izango litzateke testuen erlazio-egitura sortzeko edota prozesatzeko.



5.7 irudia: EGU06 iritzi-testua RSTz etiketatuta.

Testuaren egitura deskribatzeko, zenbait urrats bete behar dira. Hasteko, testuaren osagaiak identifikatu behar dira eta, ondoren, osagai horien arteko lotura edo erlazioa zehaztu. Testuan osagaia testu-zatia da. Testua aztertzeko testua zatitu egin behar da eta testua sortzeko testuko zatiak lotu egin behar dira. Testua aztertzean, testu-zatiak egiteari *segmentazioa* deritzo. Testu bat zatitzean, testu-zati bakoitzak proposizio bat adierazi behar du. 5.7 irudian, testua sei zatitan banatuta dago eta zati horietako bakoitzak proposizio bat adierazten du.

Ondoren, testu-zati horiek lotu egin behar dira eta hori egiteko koherentziazko erlazioak (*coherence relation*) ezartzen dira eta, testu-zatietan, koherentziazko erlazio horiek erlazio erretoriko (*rhetorical relation*) izena hartzen dute. RSTko zuhaitz-egituretan, erlazio erretorikoak errekursiboak dira. Hau da, erlazio bat beste erlazio baten unitatea da eta, modu horretan, testuko unitate guztiak zuhaitz-egituretan aurkezten dira. 5.7 irudian, testu-zatiak erlazio erretorikoz (ONDORIOA, EBALUAZIOA, etab.) lotuta daude eta erlazioen artean errekursibitatea dago. Esaterako, ONDORIOA erlazio

erretorikoa KAUSA erlazio erretorikoaren unitatea da.

5.3.2.1. Nukleartasuna

RSTn nukleartasunaren bidez erlazio-egituretan unitate bakoitzak duen garrantzia zehazten da. RSTko zuhaitz-egituretan hierarkia bat osatzen da; izan ere, diskurtso-unitate guztiek ez dute garrantzi bera. Idazleak irakurleari testuko mezua adierazi nahi dionean, unitate batzuek ideia garrantzitsuak dituzte eta beste batzuek, aldiz, bigarren mailako ideiak dira eta ideia nagusia ulertzeko baliagarriak dira. Beraz, diskurtso-unitate batzuk besteak baino garrantzitsuagoak dira. Diskurtso-unitate garrantzitsuenei nukleo-unitate (*nuclear unit*, N) deritze eta bigarren mailakoei satelite-unitate (*satellite unit*, S). Diskurtso-unitate bat nukleo- edo satelite-unitatea den ebazteari nukleartasuna (*nuclearity*) deritzo. 5.7 irudiko ONDORIOA erlazio erretorikoan, eskuineko diskurtso-unitatea ezkerrekoari lotzen zaio. Hori dela eta, ezkerreko unitatea nukleo-unitatea da eta eskuinekoa satelite-unitatea.

Nukleartasunak lotura du koherentziarekin, izan ere, idazleak testu batean adierazi nahi duena diskurtso-egiturako nukleo-unitateekin ulertzen da. Satelite-unitateekin, aldiz, testuaren ulermena zaila da, bertan dauden ideiak bigarren mailakoak direlako eta koherentziarik ez dagoelako. Erlazio nukleobakardunetan, satelitea nukleoari erlazio erretorikoz batzen zaio (N-S) eta erlazio horri *erlazio hipotaktiko* deritzo. Bi nukleoz (N-N) osatuta dauden erlazioei, aldiz, *erlazio parataktiko* deritze.

5.3.2.2. Erlazio erretorikoak

RSTn bi motatako erlazioak bereizten dira:

- Erlazio nukleoaniztunak (N-N). Mota honetako erlazioak nukleoz bakarrik osatuta daude eta, horrenbestez, erlazioa osatzen duten unitate guztiak maila berean daude eta unitateen artean ez da hierarkiarik sortzen. Erlazio nukleoaniztasunak hauek dira: KONJUNTZIOA, KONTRASTEIA, DISJUNTZIOA, BATERATZEA, LISTA, BIRFORMULAZIO NUKLEOANIZTUNA eta SEKUENTZIA.

- Erlazio nukleobakardunak (N-S/S-N). Mota honetako erlazioetan, erlazioa osatzen duten unitateak ez daude maila berean eta hierarkia bat sortzen da. Nukleo-unitateak (N) adierazi nahi denaren zati garrantzitsuena biltzen du eta satelite-unitateak (S) nukleo-unitatean aipaturiko gaia hobeto ulertzen informazioa ematen du. Erlazio nukleobakardunak esaldien eta paragrafoen barruan ager daitezke. Mann eta Thompsonen (1988) bi motatako erlazio nukleobakardunak bereizten dituzte.
 - Edukizko erlazioak (*subject matter*). Unitateen arteko loturaren efektuak izaera semantikoa duenean gertatzen da. Hots, idazleak irakurleari bi unitateen artean erlazio bera dagoela adierazi nahi dio. ZIRKUNSTANTZIA, BALDINTZA, ELABORAZIOA, EBALUAZIOA, INTERPRETAZIOA, METODOA, KAUSA, ONDORIOA, AUKERA, HELBURUA, ARAZO-SOLUZIOA, EZ-BALDINTZATZAILEA eta ALDERANTZIZKO BALDINTZA mota honetako erlazioak dira.
 - Aurkezpenezko erlazioak (*presentational*). Unitateen arteko loturak izaera erretorikoa duenean gertatzen da. Kasu honetan, lotura horrek helburutzat irakurlearengan efektu jakin bat sortzea du. ANTITESIA, TESTUINGURUA, KONTZETSIOA, AHALBIDERATZEA, EBIDENTZIA, JUSTIFIKAZIOA, MOTIBAZIOA, PRESTATZEA, BIRFORMULAZIOA eta LABURPENA aurkezpenezko erlazioak dira.

5.3.2.3. Unitate zentrala

Iruskietak (2014) aipatzen duen moduan, pertsona batek testu baten laburpena egiten duenean, testuaren koherentzia globala edo makroegitura esplizitu egiteko gaitasuna erakusten du. Laburpen zientifiko eta akademikoetan ohikoak dira “The principal aim of this paper is to investigate ...” moduko egitura edo formulak. Horiek kasurik gehienetan adierazgailuak izan ohi dituzte eta adierazgailu horiek gramatika-kategoria ezberdinetakoak izan daitezke. Lan akademikoetan, ohikoak dira gramatika-kategoria ezberdinetako adierazgailu hauek: izenak (*paper, article, method, result* eta abar), aditzak (*discuss, introduce, analy-* eta *stud-*, besteak beste), erakusleak (*this, some*

eta abar) eta azkenik, izenordainak (*we, I*). Hala ere, adierazgailu guztiek ez dute beti makroegiturako gai nagusia adierazten; izan ere, mikroegitura adierazteko ere erabil baitaitezke.

RST jarraituta eginiko zuhaitz-egitura batean, gai nagusia da erlazio-egitura-ko oinarrizko diskurtso-unitaterik (EDU) garrantzitsuena eta adierazgailuek hori identifikatzen lagun dezakete. Hau da, unitate zentrala da testuko EDU guztien artean garrantzitsuena eta zuhaitz-egitura eraiki nahi denean, bera erdigunetzat hartuta hasi behar da eraikitzen. Unitate zentralaren beste ezaugarri bat testuko diskurtso-unitate ez-errekurtsibo bakarra izatea da. Hots, unitate zentrala ez da inoiz beste erlazio baten satelite-unitatea izango.

RSTn oinarrirituta egindako zuhaitz-egituran, beti egongo da unitate zentrala, nahiz eta tesi-adierazpen edo esaldi tematikorik ez egon. Izan ere, beti egon behar du zentralitatea duen nukleo-unitate bat eta ez-errekurtsiboa dena.

5.7 irudian, unitate zentrala *Eguraldia dotoreagoa eta giro goxoagoa ere bai* testu-zatia da. *Eguraldia* hitza makroegiturako gai nagusia izateak testu-zati hori unitate zentrala dela identifikatzen laguntzen digu.

5.4 Laburpena

Kapitulu honetan, euskararen hizkuntzako maila ezberdinetan dauden testuinguruko balentzia-aldatzaileak identifikatu eta horiek hitz eta sintagmen orientazio semantikoan eta sentimendu-balentzian duten eragina neurtu dugu. Emaiza 5.8 irudian ageri da.

Fonologia	Morfologia	Sintaxia	Diskurtsoa
<ʂ>/<ʒ> -> <x> <t> -> <tt> <ʒ> -> <tx>	-garri -min -tsu -tza -en -tzar -zale -nahi -gura	ezin	· unitate zentrala · nukleoa · erlazioetako azken zatia
	-gabe des- -egi -keria ez-/ez ezin-/ezin -xe/-txe -txo -gaitz -ska/-xka -txa a-	ezin gabe izan ezik salbu ezta ez	
	-ezia -tasun	· edo/edota/ala + ez · ez ezik · egitura lexikalizatuak	

5.8 irudia: Euskarazko balentzia-aldatzaileak hizkuntza maila ezberdinetan.

Euskararen testuinguruko balentzia-aldatzaileak hizkuntza maila ezberdinetan daude eta beren eragina mota ezberdinetakoa da: hitzaren, sintagmaren edo esaldiaren sentimendu-balentzia indartu edo ahuldu egin dezakete. Halaber, badaude sentimendu-balentzian eraginik ez duten elementuak ere.

5.8 irudiko taulako lehen zutabearen, euskarazko balentzia-aldatzaile fonologikoak zerrendatuta ageri dira. Bustidura adierazkorra balentzia-aldatzaile fonologikoa da eta beti hitzaren sentimendu-balentzia indartzen du. Izan ere, bustidura adierazkorra, samurtasuna edo gertutasuna adierazteko erabiltzen da eta mezuaren hartzaileari baliabide fonologiko hori erabiliz zuzentzean intentsitate indartu egiten da.

Bigarren zutabearen, euskarazko hizkiak zerrendatuta eta sailkatuta ageri dira. Ahultzen duten balentzia-aldatzaile morfologikoak gehiago dira indar-

tzen dutenak baino. Kopuruaren ezberdintasunaren jatorrian ezeztapenezko hizkiak daude, horiek sentimendu-balentzia beti ahultzen baitute. Kontuan hartu behar da, ezeztapenezko hizkiek ± 4 sentimendu-balentzia dutela, horrenbestez, hitzaren balentziaren zeinua alda dezakete baldin eta hitzaren sentimendu-balentzia ± 3 edo baxuagoa bada.

Hirugarren zutabean, syntaxiko eta zehazki ezeztapeneko balentzia-aldatzaileak ageri dira. Ezeztapen-marka gehienek sintagmaren edo esaldiaren sentimendu-balentzia ahultzen dute. Kasu honetan ere, aintzat hartu behar da ezeztapen-marka horiek ± 4 balentzia dutela eta sintagma edo esaldiaren sentimendu-balentzia ± 3 edo txikiagoa bada, horien sentimendu-balentziaren zeinuan alderantzikatzea gertatzen da. Bestalde, *ezin* ezeztapen-markaren eskuinean adjektibo edo adberbioren bat badago, ezeztapen-marka horrek sentimendu-balentziaren indartzea eragina du. Bukatzeko, ezeztapen-markak ez du eraginik hautakariak diren lokailuekin agertzean, emendiozko lokailutzat edo ezeztapen kontrastibotzat jotzen den *ez ezik* egiturarekin agertzean eta ezeztapen-marka egitura lexikalizatu baten parte denean.

Azken zutabean, berriz, diskurtso-egiturari lotutako balentzia-aldatzaileak daude. Nukleartasunak, erlaziozko diskurtso-egituretako azken unitateak eta unitate zentralak eragina dute diskurtso-unitateek edota erlaziozko diskurtso-egiturek orientazio semantiko positiboa edo negatiboa izatean.

ONDORIOAK

Ekarpenak, mugak eta etorkizuneko lanak

Tesi-lan honen hasieran aipatu dugun moduan, gaur egun sarean iritzi-testu ugari egonda, testu horien informazio subjektiboa erauztea eta, zehazki, testu horiek balorazio positiboa edo negatiboa duten automatikoki jakitea abantaila bat da bai norbanakoentzat, bai gizartearentzat. Norabide horretan, euskarazko iritzi-testuetako informazio subjektiboa erauzteko tresna eta baliabideak garatu ditugu eta, halaber, euskarazko testuinguruko balentzia-aldatzaileak landu ditugu.

Jarraian, tesi-lan honetan euskarari dagokionez sentimenduen analisiaren inguruan egindako ekarpenak eta tesi-lanak izan dituen mugak aurkeztuko ditugu. Etorkizunera begira ditugun erronkak ere aipatuko ditugu.

6.1 Ekarpenak

Tesi hau euskarazko testu idatzien sentimenduen analisia lantzen duen lana da hizkuntzalaritza aplikatuaren ikuspegitik. Iritzi-testuak biltzen dituen corpus bat osatu eta sentimenduen lexikoi bat garatu ondoren, hitzen balentzietan eragiten duten fenomenoak aztertu ditugu. Jarduera horien ondorio dira garatu ditugun baliabideak eta tresnak.

6.1.1. Sentimenduen analisirako baliabideak eta tresnak

Euskarazko iritzi-testuen corpus etiketatua, euskarazko sentimenduen lexikoia eta dokumentu mailako sentimenduen sailkatzailea garatu ditugu.

- Euskarazko Iritzi Corpora (Alkorta *et al.*, 2016). Corpus horrek sei domeinutako 240 iritzi-testu biltzen ditu. Diskurtso-egituraren nahiz subjektibitateaz etiketatuta dago. Diskurtso-egitura etiketatzeko RST hurbilpena (Mann eta Thompson, 1988) erabili dugu eta 70 iritzi-testu etiketatu ditugu. 192 iritzi-testuren unitate zentralak ere identifikatu ditugu. Halaber, testuek orientazio semantiko positiboa edo negatiboa duten ere zehaztu dugu. Gainera, 28 iritzi-testutan, erlaziozko diskurtso-egitura mota batzuei eta beren osagaiei (nukleoa eta satelitea) orientazio semantikoa (positiboa edo negatiboa den) nahiz sentimendu-balentzia (zenbakizko sentimenduen balioa) esleitu diegu.
- *Sentitegi* izeneko euskarazko sentimenduen lexikoia (Alkorta *et al.*, 2018). Sentimenduen lexikoi horrek 5 gramatika-kategoritako 1.237 hitz biltzen ditu. Bere oinarrian SO-CAL tresnaren gaztelaniazko lexikoia (Brooke *et al.*, 2009) dago eta ingelesezko lexikoz (Taboada *et al.*, 2011) ere aberastuta dago. Sentimenduen lexikoia sei domeinuetarako egokitu dugu (eguraldia, kirola, politika, musika, zinema eta literatura) eta, horretarako, Euskarazko Iritzi Corpora (Alkorta *et al.*, 2016) baliatu dugu.
- Tesi-lan honen beste ekarpen bat dokumentu mailako eta lexikoian oinarritutako sentimenduen sailkatzailea da. Euskarazko sentimenduen sailkatzailea garatzeko SO-CAL tresnan (Taboada *et al.*, 2011) oinarritu gara.

6.1.2. Euskarazko balentzia-aldatzaileen azterketa

Sentimenduen analisisia lantzeko, hitzen sentimendu-balentzian eragiten duten hizkuntza maila ezberdinetako testuinguruko balentzia-aldatzaileak aztertu

ditugu. Balentzia-aldatzaileak aztertzerakoan, Polanyi eta Zaenenen (2006) lanean oinarritu gara.

- Fonologia eta morfologia mailako balentzia-aldatzaileak aztertu ditugu. Bertan ikusi dugunez, fonologian bustidura adierazkorra izeneko fenomeno dago eta beren eragina da hitzaren sentimendu-balentzia indartzea. Morfologian, berriz, atzizkiak (adibidez, *-txo/-tzar*) aurkitu ditugu hitzaren balentzia indartu edo ahuldu dezaketenak. Gainera, ezeztapenarekin loturiko zenbait aurrizki (adibidez, *des-*) eta atzizki (adibidez, *-ezin*) ere aurkitu ditugu.
- Maila sintaktikoa ere aztertu dugu eta, zehazki, ezeztapen-markak eta irismena. Azterketak erakusten duenez, ezeztapen-markek hitzaren edo sintagmaren balentzia indartu (adibidez, *ezin*) edo ahuldu (esaterako, *gabe*) dezakete eta badaude aldaketarik eragiten ez duten kasuak (esaterako, *baino ez*) ere.

Halaber, aipatzekoa da ezeztapen-marka batek (*ezin*) jokamolde bi-koitza duela inguruan duen hitzaren gramatika-kategoriaren arabera. Ezeztapen-marka batzuk beste hitz jakin batzuekin maiztasun handiz agertzen direla ere erakutsi digute emaitzek. Egitura lexikalizatuak deitu diegu hauei. Irismenari dagokionez, *Murritzapen Gramatika* (Karlsson *et al.*, 1995) hurbilpena erabiliz, berau ezeztapen-markak eta egitura lexikalizatuak identifikatzea baino zailagoa dela behatu dugu, egitura eta luzera irregularrekoa baita.

- Azkenik, diskurtso maila ere landu dugu eta, horretarako, RST hurbilpena (Mann eta Thompson, 1988) erabili dugu. Lanean, erlaziozko diskurtso-egiturei eta unitate zentralari eman diogu garrantzia. Erlaziozko diskurtso-egituretan, ikusi dugu erlazioetako nukleoak edo azken EDUak osagai horien beraren eta erlaziozko diskurtso-egituraren arteko orientazio semantikoaren adostasuna indartzen dutela. Horretaz gain, emaitzek erakusten dute erlaziozko diskurtso-egitura bat unitate zentraletik gertuago egon ahala, erlazioaren beraren eta testu osoaren orientazio semantikoaren adostasuna indartu egiten dela. Horrenbes-

tez, emaitzek erakusten dute erlazioetan nukleoa eta azken EDUa eta testuan unitate zentrala balentzia indartzaile direla.

Unitate zentralean, balentziadun hitzak nola agertzen diren ere aztertu dugu. Emaitzen arabera, unitate zentraletako hitzen % 12,40k balentzia dute, batez ere, adjektibo, izen eta aditzek. Unitate zentralean, balentziadun hitzetan domeinuen arabera ezberdintasunak daudela behatu dugu. Esaterako, kirolaren domeinuan aditz-kategoriako hitzak dira balentziadunak, batik bat. Eguraldiaren domeinuan, ordea, adjektibo-kategoriako hitzak dira balentziadunak, batez ere.

6.2 Lanaren mugak

Atal honetan, tesi-lan honek izan dituen mugak aipatuko ditugu:

- Euskaran dauden testuinguruko balentzia-aldatzaile gehiago identifikatzen eta aztertzen jarraituko dugu. Tesi-lan honetan, hizkuntza maila bakoitzeko balentzia-aldatzaile bat aztertu dugu, baina oraindik badira aztertzeko geratu direnak. Hori da lan honek izan duen mugetariko bat. Aztertu gabe gelditu diren balentzia-aldatzaileen artean, baldintza, galde-perpauzak edota konplexuagoa den ironia aipa ditzakegu. Balentzia-aldatzaileetako batzuek euskaran berezitasunak eduki ditzakete eta beste batzuk, aldiz, hizkuntza ezberdinetan antzekoak izan.
- Diskurtso mailako azterketa gehiago sakondu nahi dugu. Lan honetan, erlaziozko diskurtso-egiturak eta unitate zentrala aztertu ditugu. Baina, sentimenduen analisisian eragin dezaketen diskurtso-egituraren elementu gehiago egon daitezke. Horregatik, etorkizunean, haratago joan nahi dugu eta diskurtso-egiturako beste osagai batzuk ere aztertu nahi ditugu. Beste modu batean esanda, diskurtso-egituraren dauden beste elementu batzuk aztertu nahi ditugu, azpiosagai zentrala esaterako. Halaber, erlaziozko diskurtso-egitura motetan ere sakondu nahi dugu eta beren artean egon daitezkeen ezberdintasunak aurkitu nahi ditugu.

6.3 Etorkizuneko lanak

Sentimenduen analisiaren azterketan eta beraren tratamendu konputazionalan hurrengo lan-ildoak aurreikusten dugu etorkizunerako:

- Euskarazko Iritzi Corpora (Alkorta *et al.*, 2016) are anitzagoa izatea. Gaur-gaurkoz corpusak egunkari eta webgune espezializatutako 6 domeinutako iritzi-testuak biltzen ditu. Etorkizuneko asmoa jendeak idazten dituen iritzi-testuak biltzea da. Metodologiaren garapenean aipatu bezala, jende arruntak idatzitako iritzi-testu gutxi aurkitu ditugu edota aurkitu ditugunen kalitatea ez da egokia izan diskurtsoa lantzeko. Asmoa da mota horretako testuak biltzea, oraindik aztertu gabe dauden bestelako fenomenoak ager baitaitezke. Adibidez, jende arruntak idatzita testuetan posible da letra bat errepikatuz hitzak luzatzea, baita emotikonoak bezalako elementuak agertzea ere, eta horiek orientazio semantikoan ere eragina izan dezakete.
- Corpusaren diskurtso-egituraren etiketatzea hobetu eta zabaldu. Mementoz, corpuseko 240 testuetatik 70 daude etiketatuta eta kopuru hori handitu egin nahi dugu.
- *Sentitegi* sentimenduen lexikoa (Alkorta *et al.*, 2018) handitzea eta metodologian izandako gertaerei aurre egitea. Ikusi dugun moduan, sortu dugun lexikoia kalitatea ona da, baina oraindik ez da gauza subjektibitatea adierazten duten hitz edota esamolde guztiak identifikatzeko. Ondorioz, bere tamaina handitu nahi dugu eta domeinu gehiagotarako erabilgarria izatea lortu. Bukatzeko, hitz polisemikoen arazoa ere ebazti nahi dugu, hau da, testuinguru ezberdinetan, aurkako bi orientazio dituzten hitzekin zer egin erabaki. Izan ere, *ikaragarria* bezalako hitzak aurkako orientazio semantikoa du, esaterako, pelikula bati buruz edo istripu bati buruz ari garenean.
- Dokumentu mailako sentimenduen sailkatzailea gehiago garatu nahi dugu. Tresna, mementoz euskarazko lexikoiaz, lematizatzaileaz eta erregela orokorrez osatuta dago, baina bertan aspektu gehiago integratu nahi ditugu. Esaterako, tesi-lan honetan landutako balentzia-

aldatzailei lotutako modulua eratu nahiko genuke eta, modu horretan, tresnaren kalitatea igotzen den aztertu beharko litzateke. Modulu horretan, fonologiatik hasi eta diskurtso-egiturarainoko balentzia-aldatzailei buruzko erregelak garatu nahiko genituzke.

Bibliografía

- Aduriz I., Aldezabal I., Alegria I., Arriola J., Díaz de Ilarraza A., Ezeiza N., eta Gojenola K. Finite state applications for Basque. *EACL'2003 Workshop on Finite-State Methods in Natural Language Processing*, 3–11, 2003.
- Aduriz I., Aranzabe M.J., Arriola J.M., Díaz de Ilarraza A., Gojenola K., Oronoz M., eta Uria L. A Cascaded Syntactic Analyser for Basque. In Gelbukh A., editor, *Computational Linguistics and Intelligent Text Processing*, 124–134, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. ISBN 978-3-540-24630-5.
- Aduriz I., Arriola J.M., Artola X., Díaz de Ilarraza A., Gojenola K., eta Maritxalar M. Morphosyntactic disambiguation for Basque based on the Constraint Grammar Formalism. *Proceedings of Recent Advances in NLP (RANLP97), Tzigov Chark (Bulgary)*, 282–288, 1997.
- Agerri R., Bermudez J., eta Rigau G. IXA pipeline: Efficient and Ready to Use Multilingual NLP tools. In Chair) N.C.C., Choukri K., Declerck T., Loftsson H., Maegaard B., Mariani J., Moreno A., Odijk J., eta Piperidis S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.
- Aldezabal I., Ansa O., Arrieta B., Artola X., Ezeiza A., Hernández G., eta Lersundi M. EDBL: A general lexical basis for the automatic processing of

BIBLIOGRAFIA

- Basque. *Proceedings of the IRCS Workshop on linguistic databases*. IRCS Workshop on linguistic databases., 2006.
- Alegria I., Aranzabe M., Ezeiza A., Ezeiza N., eta Urizar R. Robustness and customisation in an analyser/lemmatiser for Basque. *LREC-2002 Customizing knowledge in NLP applications workshop*, 1–6, 2002.
- Alistair K. eta Diana I. Sentiment classification of movie and product reviews using contextual valence shifters. *Proceedings of FINEXIN*, 2005.
- Alkorta J., Gojenola K., eta Iruskieta M. Creating and evaluating a polarity-balanced corpus for Basque sentiment analysis. *IWoDA16 Fourth International Workshop on Discourse Analysis. Santiago de Compostela, September, 29 lib.*, 2016.
- Alkorta J., Gojenola K., eta Iruskieta M. SentiTegi: Semi-manually Created Semantic Oriented Basque Lexicon for Sentiment Analysis. *Computación y Sistemas*, 22(4), 2018.
- Altuna B., Aranzabe M.J., eta Díaz de Ilarraza A. Euskarazko ezeztapenaren tratamendu automatikorako azterketa. *Iñaki Alegria, Ainhoa Latatu, Miren Josu Ormaetxebarria eta Patri Salaberri (arg.), II. IkerGazte, Nazioarteko Ikerketa Euskaraz: Giza Zientziak eta Arteak, 127-134, Udako Euskal Unibertsitatea (UEU), Bilbo*, 2017.
- Altuna P., Salaburu P., Goenaga P., Lasarte M.P., Akesolo L., Azkarate M., Charriton P., Eguskitza A., Haritschelhar J., King A., Larrarte J.M., Mujika J.A., Oyharçabal B.n., eta Rotaetxe K. Euskal Gramatika Lehen urratsak (EGLU) II. *Euskaltzaindiko Gramatika batzordea, Euskaltzaindia, Bilbo*, 1985.
- Banea C., Mihalcea R., Wiebe J., eta Hassan S. Multilingual Subjectivity Analysis Using Machine Translation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, 127–135, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1613715.1613734>.

- Barnes J., Badia T., eta Lambert P. MultiBooked: A Corpus of Basque and Catalan Hotel Reviews Annotated for Aspect-level Sentiment Classification. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, Miyazaki, Japan, May 2018a. European Languages Resources Association (ELRA). URL <https://www.aclweb.org/anthology/L18-1104>.
- Barnes J., Klinger R., eta Schulte im Walde S. Bilingual Sentiment Embeddings: Joint Projection of Sentiment Across Languages. *CoRR*, abs/1805.09016, 2018b. URL <http://arxiv.org/abs/1805.09016>.
- Bautin M., Vijayarenu L., eta Skiena S. International sentiment analysis for news and blogs. *ICWSM*, 19–26, 2008.
- Beigman Klebanov B., Burstein J., Madnani N., Faulkner A., eta Tetreault J. Building subjectivity lexicon(s) from scratch for essay data. *Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing - Volume Part I, CICLing'12*, 591–602, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-28603-2. URL http://dx.doi.org/10.1007/978-3-642-28604-9_48.
- Benesty J., Chen J., Huang Y., eta Cohen I. Pearson correlation coefficient. *Noise reduction in speech processing*, 1–4. Springer, 2009.
- Blair-Goldensohn S., Hannan K., McDonald R., Neylon T., Reis G., eta Reynar J. Building a Sentiment Summarizer for Local Service Reviews. *WWW Workshop on NLP Challenges in the Information Explosion Era (NLPIX)*, 2008.
- Blitzer J., Dredze M., eta Pereira F. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 440–447, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/P07-1056>.

BIBLIOGRAFIA

- Boldrini E., Balahur A., Martínez-Barco P., eta Montoyo A. EmotiBlog: a finer-grained and more precise learning of subjectivity expression models. *Proceedings of the Fourth Linguistic Annotation Workshop*, 1–10. Association for Computational Linguistics, 2010.
- Bond F., Ohkuma T., Da Costa L.M., Miura Y., Chen R., Kuribayashi T., eta Wang W. A Multilingual Sentiment Corpus for Chinese, English and Japanese. *Emotion and Sentiment Analysis*, page 59, 2016.
- Boucher J. eta Osgood C.E. The pollyanna hypothesis. *Journal of verbal learning and verbal behavior*, 8(1):1–8, 1969.
- Brooke J., Tofiloski M., eta Taboada M. Cross-linguistic sentiment analysis: From English to Spanish. *Proceedings of the International Conference RANLP-2009*, 50–54, Borovets, Bulgaria, September 2009. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/R09-1010>.
- Cambria E., Das D., Bandyopadhyay S., eta Feraco A. *A Practical Guide to Sentiment Analysis*, 5 lib. 01 2017. ISBN 978-3-319-55392-4.
- Carenini G., Ng R., eta Pauls A. Multi-Document Summarization of Evaluative Text. *11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, April 2006. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/E06-1039>.
- Chen Y. eta Skiena S. Building Sentiment Lexicons for All Major Languages. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 383–389, Baltimore, Maryland, June 2014. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/P14-2063>.
- Clematide S., Gindl S., Klenner M., Petrakis S., Remus R., Ruppenhofer J., Waltinger U., eta Wiegand M. MLSA - A Multi-layered Reference Corpus for German Sentiment Analysis. In Calzolari N., Choukri K., Declerck T., Dogan M.U., Maegaard B., Mariani J., Odijk J., eta Piperidis S., editors,

-
- Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey, May 23-25, 2012*, 3551–3556, 2012.
- Cohen J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.
- Cruz F.L., Troyano J.A., Pontes B., eta Ortega F.J. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, 41(13):5984–5994, 2014.
- Dadvar M., Hauff C., eta de Jong F. Scope of negation detection in sentiment analysis. *Proceedings of the Dutch-Belgian Information Retrieval Workshop, DIR 2011*, 16–20. University of Amsterdam, 2 2011. ISBN not assigned.
- Das D. eta Taboada M. RST Signalling Corpus: A Corpus of Signals of Coherence Relations. *Lang. Resour. Eval.*, 52(1):149–184, March 2018. ISSN 1574-020X. URL <https://doi.org/10.1007/s10579-017-9383-x>.
- Das D. eta Martins A.F. A survey on automatic text summarization. *Literature Survey for the Language and Statistics II course at CMU*, 4(192-195): 57, 2007.
- Ding X., Liu B., eta Yu P.S. A Holistic Lexicon-based Approach to Opinion Mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08*, 231–240, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-927-2. URL <http://doi.acm.org/10.1145/1341531.1341561>.
- Ding X., Liu B., eta Zhang L. Entity discovery and assignment for opinion mining applications. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, 1125–1134, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-495-9. URL <http://doi.acm.org/10.1145/1557019.1557141>.

BIBLIOGRAFIA

- Du W., Tan S., Cheng X., eta Yun X. Adapting Information Bottleneck Method for Automatic Construction of Domain-oriented Sentiment Lexicon. *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, 111–120, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-889-6. URL <http://doi.acm.org/10.1145/1718487.1718502>.
- Edmonds P. eta Hirst G. Near-synonymy and Lexical Choice. *Comput. Linguist.*, 28(2):105–144, June 2002. ISSN 0891-2017. URL <http://dx.doi.org/10.1162/089120102760173625>.
- Egaña I. *Kritikarako hurbilketa literaturaren soziologiatik. Egunkari eta aldizkarietako euskal literatur kritikaren analisisa (1975-2005)*. Doktoretzatesia, Universidad del País Vasco-Euskal Herriko Unibertsitatea, 2013.
- Egg M. eta Redeker G. How Complex is Discourse Structure? *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Languages Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2010/pdf/796_Paper.pdf.
- Eguchi K. eta Lavrenko V. Sentiment Retrieval Using Generative Models. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, 345–354, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. ISBN 1-932432-73-6. URL <http://dl.acm.org/citation.cfm?id=1610075.1610124>.
- Elhuyar H.Z. Elhuyar hiztegia: euskara-gaztelania, castellano-vasco. Usurbil: Elhuyar, 2013.
- Euskaltzaindia. Euskal Gramatika. Lehen urratsak I, Iruñea: Euskaltzaindia, 1985.
- Euskara Institutua E. Sareko euskal gramatika, 2011. URL www.ehu.es/seg.

- Fernández J., Boldrini E., Gómez J.M., eta Martínez-Barco P. Análisis de sentimientos y minería de opiniones: el corpus emotiblog. *Procesamiento del lenguaje natural*, 47:179–187, 2011.
- Fundazioa K.H. *Euskal hiztegi entziklopedikoa*. Klaudio Harluxet Fundazioa, 1995.
- Gamon M. Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis. *Proceedings of the 20th International Conference on Computational Linguistics, COLING '04*, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. URL <https://doi.org/10.3115/1220355.1220476>.
- Gamon M., Aue A., Corston-Oliver S., eta Ringger E. Pulse: Mining Customer Opinions from Free Text. *Proceedings of the 6th International Conference on Advances in Intelligent Data Analysis, IDA'05*, 121–132, Berlin, Heidelberg, 2005. Springer-Verlag. ISBN 3-540-28795-7, 978-3-540-28795-7. URL http://dx.doi.org/10.1007/11552253_12.
- Ghosh A., Li G., Veale T., Rosso P., Shutova E., Barnden J., eta Reyes A. SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 470–478, Denver, Colorado, jun 2015. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/S15-2080>.
- Guo H., Zhu H., Guo Z., Zhang X., eta Su Z. Opinionit: A text mining system for cross-lingual opinion analysis. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10*, 1199–1208, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0099-5. URL <http://doi.acm.org/10.1145/1871437.1871589>.
- Halliday M.A.K., Matthiessen C., eta Halliday M. *An introduction to functional grammar*. Routledge, 2014.
- Hassan A., Abu-Jbara A., Lu W., eta Radev D. A random walk: Based model for identifying semantic orientation. *Comput. Linguist.*, 40(3):539–

BIBLIOGRAFIA

- 562, September 2014. ISSN 0891-2017. URL http://dx.doi.org/10.1162/COLI_a_00192.
- Hatzivassiloglou V. eta McKeown K.R. Predicting the Semantic Orientation of Adjectives. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, ACL '98/EACL '98*, 174–181, Stroudsburg, PA, USA, 1997. Association for Computational Linguistics. URL <https://doi.org/10.3115/976909.979640>.
- Hu M. eta Liu B. Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04*, 168–177, New York, NY, USA, 2004. ACM. ISBN 1-58113-888-1. URL <http://doi.acm.org/10.1145/1014052.1014073>.
- Hu Y.H., Chen Y.L., eta Chou H.L. Opinion Mining from Online Hotel Reviews A Text Summarization Approach. *Inf. Process. Manage.*, 53(2): 436–449, March 2017. ISSN 0306-4573. URL <https://doi.org/10.1016/j.ipm.2016.12.002>.
- Iruskieta M. Pragmatikako erlaziozko diskurtso-egitura: deskribapena eta bere ebaluazioa hizkuntzalaritza konputazionalen (a description of pragmatics rhetorical structure and its evaluation in computational linguistic). *Doktore-tesia. EHU, Informatika Fakultatea*, 2014.
- Iruskieta M., Aranzabe M.J., Díaz de Ilarraza A., Gonzalez I., Lersundi M., eta Lopez de Lacalle O. The RST Basque TreeBank: an online search interface to check rhetorical relations. *4th workshop RST and discourse studies*, 40–49, 2013.
- Iruskieta M., da Cunha I., eta Taboada M. Principles of a qualitative method for rhetorical analysis evaluation: A contrastive analysis English-Spanish-Basque. *Language Resources and Evaluation*, 49(2):263–309, 2015.
- Jia L., Yu C., eta Meng W. The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness. *Proceedings of the 18th ACM Confe-*

-
- rence on Information and Knowledge Management*, CIKM '09, 1827–1830, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-512-3. URL <http://doi.acm.org/10.1145/1645953.1646241>.
- Jindal N. et al Liu B. Mining Comparative Sentences and Relations. *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, 1331–1336. AAAI Press, 2006. ISBN 978-1-57735-281-5. URL <http://dl.acm.org/citation.cfm?id=1597348.1597400>.
- Kanayama H. et al Nasukawa T. Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, 355–363, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. ISBN 1-932432-73-6. URL <http://dl.acm.org/citation.cfm?id=1610075.1610125>.
- Karlsson F., Voutilainen A., Heikkilä J., et al Anttila A., editors. *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Walter de Gruyter & Co., Hawthorne, NJ, USA, 1995. ISBN 3110141795.
- Karoui J., Farah B., Moriceau V., Patti V., Bosco C., et al Aussenac-Gilles N. Exploring the Impact of Pragmatic Phenomena on Irony Detection in Tweets: A Multilingual Corpus Study. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 262–272. Association for Computational Linguistics, 2017. URL <http://aclweb.org/anthology/E17-1025>.
- Kennedy A. et al Inkpen D. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*, 22:110–125, 05 2006.
- Kim J., Li J.J., et al Lee J.H. Evaluating Multilanguage-comparability of Subjectivity Analysis Systems. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, 595–603, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1858681.1858742>.

BIBLIOGRAFIA

- Kim S.M. eta Hovy E. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, SST '06, 1–8, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. ISBN 1-932432-75-2. URL <http://dl.acm.org/citation.cfm?id=1654641.1654642>.
- Kim S.M., Pantel P., Chklovski T., eta Pennacchiotti M. Automatically assessing review helpfulness. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, 423–430, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. ISBN 1-932432-73-6. URL <http://dl.acm.org/citation.cfm?id=1610075.1610135>.
- Kouloumpis E., Wilson T., eta Moore J. Twitter sentiment analysis: The good the bad and the omg! *ICWSM*, 11:538–541, 2011.
- Landis J.R. eta Koch G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 1977.
- Leturia I., Arregi X., eta Sarasola K. Web a euskarazko corpus gisa. *Ekaia*, 27:281, 12 2014.
- Li F., Huang M., Yang Y., eta Zhu X. Learning to Identify Review Spam. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*, IJCAI'11, 2488–2493. AAAI Press, 2011. ISBN 978-1-57735-515-1. URL <http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-414>.
- Li H., Cheng X., Adson K., Kirshboim T., eta Xu F. Annotating opinions in German political news. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, 1183–1188, Istanbul, Turkey, May 2012. European Languages Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2012/pdf/640_Paper.pdf.
- Lim E.P., Nguyen V.A., Jindal N., Liu B., eta Lauw H.W. Detecting Product Review Spammers Using Rating Behaviors. *Proceedings of the 19th*

-
- ACM International Conference on Information and Knowledge Management*, CIKM '10, 939–948, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0099-5. URL <http://doi.acm.org/10.1145/1871437.1871557>.
- Liu B. Sentiment Analysis and Subjectivity. *Handbook of natural language processing*, 2:627–666, 2010.
- Liu B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.
- Liu B., Hu M., eta Cheng J. Opinion Observer: Analyzing and Comparing Opinions on the Web. *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, 342–351, New York, NY, USA, 2005. ACM. ISBN 1-59593-046-9. URL <http://doi.acm.org/10.1145/1060745.1060797>.
- Liu F., Li B., eta Liu Y. Finding opinionated blogs using statistical classifiers and lexical features. *Third International AAAI Conference on Weblogs and Social Media*, 2009.
- Liu F., Wang D., Li B., eta Liu Y. Improving Blog Polarity Classification via Topic Analysis and Adaptive Methods. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 309–312, Los Angeles, California, June 2010. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/N10-1042>.
- Long C., Zhang J., eta Zhut X. A Review Selection Approach for Accurate Feature Rating Estimation. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, 766–774, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1944566.1944654>.
- Lu Y., Tsaparas P., Ntoulas A., eta Polanyi L. Exploiting Social Context for Review Quality Prediction. *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, 691–700, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8. URL <http://doi.acm.org/10.1145/1772690.1772761>.

BIBLIOGRAFIA

- Mann W.C. eta Thompson S.A. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988.
- McDonald R., Hannan K., Neylon T., Wells M., eta Reynar J. Structured Models for Fine-to-Coarse Sentiment Analysis. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 432–439. Association for Computational Linguistics, 2007. URL <http://aclweb.org/anthology/P07-1055>.
- Meng X., Wei F., Xu G., Zhang L., Liu X., Zhou M., eta Wang H. Lost in Translations? Building Sentiment Lexicons using Context Based Machine Translation. *Proceedings of COLING 2012: Posters*, 829–838, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL <https://www.aclweb.org/anthology/C12-2081>.
- Miao Y., Su J., Liu S., eta Wu K. *SO-CAL Based Method for Chinese Sentiment Analysis*, 345–351. 12 2013.
- Miller G.A. WordNet: A Lexical Database for English. *Commun. ACM*, 38(11):39–41, November 1995. ISSN 0001-0782. URL <http://doi.acm.org/10.1145/219717.219748>.
- Mohammad S.M. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion measurement*, 201–237. Elsevier, 2016.
- Mohammad S.M., Salameh M., eta Kiritchenko S. How Translation Alters Sentiment. *J. Artif. Int. Res.*, 55(1):95–130, January 2016. ISSN 1076-9757. URL <http://dl.acm.org/citation.cfm?id=3013558.3013562>.
- Morsy S.A. eta Rafea A. Improving Document-level Sentiment Classification Using Contextual Valence Shifters. *Proceedings of the 17th International Conference on Applications of Natural Language Processing and Information Systems, NLDB'12*, 253–258, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-31177-2. URL http://dx.doi.org/10.1007/978-3-642-31178-9_30.

- Mujika L.M. Morfología de la composición lexical euskérica. *Fontes linguae vasconum: Studia et documenta*, 14(39):233–272, 1982.
- Mullen T. eta Collier N. Sentiment analysis using support vector machines with diverse information sources. *In Proceedings of Conference on Empirical Methods in Natural Language Processing*, 2004.
- Na S.H., Lee Y., Nam S.H., eta Lee J.H. Improving Opinion Retrieval Based on Query-Specific Sentiment Lexicon. In Boughanem M., Berrut C., Mothe J., eta Soule-Dupuy C., editors, *Advances in Information Retrieval*, 734–738, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-00958-7.
- Nakagawa T., Inui K., eta Kurohashi S. Dependency Tree-based Sentiment Classification Using CRFs with Hidden Variables. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, 786–794, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. ISBN 1-932432-65-5. URL <http://dl.acm.org/citation.cfm?id=1857999.1858119>.
- Ngoc Phu V. eta Thi Tuoi P. Sentiment classification using Enhanced Contextual Valence Shifters. 224–229, 10 2014.
- O'Donnell M. RSTTool 2.4: A Markup Tool for Rhetorical Structure Theory. *Proceedings of the First International Conference on Natural Language Generation - Volume 14*, INLG '00, 253–256, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics. ISBN 965-90296-0-8. URL <https://doi.org/10.3115/1118253.1118290>.
- Oñederra L. *Euskal fonologia: palatalizazioa: asimilazioa eta hots sinbolismoa*. Servicio Editorial de la Universidad del País Vasco= Euskal Herriko Unibertsitatea, 1990.
- Oronoz M. *Euskarazko errore sintaktikoak detektatzeko eta zuzentzeko baliabideen garapena: datak, postposizio-lokuzioak eta komunztadura*. Doktoretzatesia, Universidad del País Vasco-Euskal Herriko Unibertsitatea, 2009.

BIBLIOGRAFIA

- Otegi A., Imaz O., Díaz de Ilarraza A., Iruskieta M., eta Uria L. ANALHITZA: a tool to extract linguistic information from large corpora in Humanities research. *Procesamiento del Lenguaje Natural*, (58):77–84, 2017.
- Pang B. eta Lee L. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05*, 115–124, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics. URL <https://doi.org/10.3115/1219840.1219855>.
- Pang B. eta Lee L. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January 2008. ISSN 1554-0669. URL <http://dx.doi.org/10.1561/1500000011>.
- Pang B., Lee L., eta Vaithyanathan S. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, EMNLP '02*, 79–86, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. URL <https://doi.org/10.3115/1118693.1118704>.
- Polanyi L. eta Zaenen A. Contextual valence shifters. 20 lib., 1–10. 01 2006.
- Qu L., Ifrim G., eta Weikum G. The Bag-of-opinions Method for Review Rating Prediction from Sparse Text Patterns. *Proceedings of the 23rd International Conference on Computational Linguistics, COLING '10*, 913–921, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1873781.1873884>.
- Quan C. eta Ren F. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, 1446–1454, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. ISBN 978-1-932432-63-3. URL <http://dl.acm.org/citation.cfm?id=1699648.1699691>.
- Rao D. eta Ravichandran D. Semi-supervised Polarity Lexicon Induction. *Proceedings of the 12th Conference of the European Chapter of the As-*

-
- sociation for Computational Linguistics*, EACL '09, 675–682, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1609067.1609142>.
- Refaee E. et al Rieser V. An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis. In Chair) N.C.C., Choukri K., Declerck T., Loftsson H., Maegaard B., Mariani J., Moreno A., Odijk J., et al Piperidis S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.
- Riloff E., Patwardhan S., et al Wiebe J. Feature subsumption for opinion analysis. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, 440–448, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. ISBN 1-932432-73-6. URL <http://dl.acm.org/citation.cfm?id=1610075.1610137>.
- Rodriguez I., Martínez-Otzeta J.M., Lazkano E., et al Ruiz T. Adaptive Emotional Chatting Behavior to Increase the Sociability of Robots. *International Conference on Social Robotics*, 666–675. Springer, 2017.
- Rushdi-Saleh M., Martínez-Valdivia M.T., Ureña; a López L.A., et al Perea-Ortega J.M. OCA: Opinion Corpus for Arabic. *J. Am. Soc. Inf. Sci. Technol.*, 62(10):2045–2054, October 2011. ISSN 1532-2882. URL <http://dx.doi.org/10.1002/asi.21598>.
- San Vicente I., Saralegi X., et al Agerri R. EliXa: A modular and flexible ABSA platform. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 748–752, Denver, Colorado, June 2015. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/S15-2127>.
- Saralegi X., San Vicente I., et al Ugarteburu I. Cross-Lingual Projections vs. Corpora Extracted Subjectivity Lexicons for Less-resourced Languages. *Proceedings of the 14th International Conference on Computational Linguistics and Intelligent Text Processing - Volume 2*, CICLing'13, 96–108,

BIBLIOGRAFIA

- Berlin, Heidelberg, 2013. Springer-Verlag. ISBN 978-3-642-37255-1. URL http://dx.doi.org/10.1007/978-3-642-37256-8_9.
- Sarasola I. *Zehazki: gaztelania-euskara hiztegia*. Alberdania, 2005.
- Sauri R. *A Factuality Profiler for Eventualities in Text*. Doktoretza-tesia, Waltham, MA, USA, 2008. AAI3304029.
- Schulz J.M., Womser-Hacker C., eta Mandl T. Multilingual Corpus Development for Opinion Mining. In Calzolari N., Choukri K., Maegaard B., Mariani J., Odijk J., Piperidis S., Rosner M., eta Tapias D., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010. European Language Resources Association (ELRA). ISBN 2-9517408-6-7.
- Shin H., Kim M., Jang H., eta Cattle A. Annotation Scheme for Constructing Sentiment Corpus in Korean. *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation*, 181–190. Faculty of Computer Science, Universitas Indonesia, 2012. URL <http://aclweb.org/anthology/Y12-1019>.
- Silvennoinen O.O. Not only apples but also oranges: Contrastive negation and register. *Studies in Variation, Contacts and Change in English*, 19, 2017.
- Snyder B. eta Barzilay R. Multiple Aspect Ranking Using the Good Grief Algorithm. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, 300–307, Rochester, New York, April 2007. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/N07-1038>.
- Sorby S. Translating news from English to Chinese: Complimentary and derogatory language usage. *Babel*, 54:19–35, 01 2008.
- Stone P.J., Dunphy D.C., Smith M.S., eta Ogilvie D.M. *The General Inquirer: A Computer Approach to Content Analysis*. 1966.

- Stone P.J. eta Hunt E.B. A Computer Approach to Content Analysis: Studies Using the General Inquirer System. *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference, AFIPS '63* (Spring), 241–256, New York, NY, USA, 1963. ACM. URL <http://doi.acm.org/10.1145/1461551.1461583>.
- Szmrecsanyi B. *Analyticity and syntheticity in the history of English*. Oxford University Press, 2012.
- Taboada M. SFU Review Corpus [Corpus]. Vancouver: Simon Fraser University, 2008.
- Taboada M., Brooke J., Tofiloski M., Voll K., eta Stede M. Lexicon-based Methods for Sentiment Analysis. *Comput. Linguist.*, 37(2):267–307, June 2011. ISSN 0891-2017. URL http://dx.doi.org/10.1162/COLI_a_00049.
- Tan P.N., Steinbach M., eta Kumar V. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005. ISBN 0321321367.
- Tsur O. eta Rappoport A. RevRank: A Fully Unsupervised Algorithm for Selecting the Most Helpful Book Reviews, 2009. URL <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/180>.
- Turney P.D. Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, 417–424, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. URL <https://doi.org/10.3115/1073083.1073153>.
- Turney P.D. eta Littman M.L. Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus. *CoRR*, cs.LG/0212012, 2002. URL <http://arxiv.org/abs/cs.LG/0212012>.
- Turney P.D. eta Littman M.L. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *ACM Trans. Inf. Syst.*, 21(4):

BIBLIOGRAFIA

- 315–346, October 2003. ISSN 1046-8188. URL <http://doi.acm.org/10.1145/944012.944013>.
- Vicente I.S. et al Saralegi X. Polarity lexicon building: to what extent is the manual effort worth? In Chair) N.C.C., Choukri K., Declerck T., Goggi S., Grobelnik M., Maegaard B., Mariani J., Mazo H., Moreno A., Odijk J., et al Piperidis S., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may 2016. European Language Resources Association (ELRA). ISBN 978-2-9517408-9-1.
- Vilares D., Garcia M., Alonso M.A., et al Gómez-Rodríguez C. Towards Syntactic Iberian Polarity Classification. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 67–73. Association for Computational Linguistics, 2017. URL <http://aclweb.org/anthology/W17-5209>.
- Villarroel Ordenes F., Ludwig S., De Ruyter K., Grewal D., et al Wetzels M. Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research*, 43(6):875–894, 2017.
- Vinodhini G. et al Chandrasekaran R. Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292, 2012.
- Wan X. Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, 553–561, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1613715.1613783>.
- Wan X. Co-training for Cross-lingual Sentiment Classification. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1, ACL '09*, 235–243, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. ISBN 978-1-932432-45-9. URL <http://dl.acm.org/citation.cfm?id=1687878.1687913>.

- Wang G., Xie S., Liu B., eta Yu P.S. Identify Online Store Review Spammers via Social Review Graph. *ACM Trans. Intell. Syst. Technol.*, 3(4):61:1–61:21, September 2012. ISSN 2157-6904. URL <http://doi.acm.org/10.1145/2337542.2337546>.
- Wei B. eta Pal C. Cross Lingual Adaptation: An Experiment on Sentiment Classifications. *Proceedings of the ACL 2010 Conference Short Papers*, ACLShort '10, 258–262, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1858842.1858890>.
- Westerski A. Sentiment Analysis: Introduction and the State of the Art overview. *Universidad Politecnica de Madrid, Spain*, 211–218, 2007.
- Wiebe J. Learning Subjective Adjectives from Corpora. *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, 735–740. AAAI Press, 2000. ISBN 0-262-51112-6. URL <http://dl.acm.org/citation.cfm?id=647288.721121>.
- Wiebe J. eta Mihalcea R. Word Sense and Subjectivity. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, ACL-44, 1065–1072, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. URL <https://doi.org/10.3115/1220175.1220309>.
- Wiebe J., Wilson T., Bruce R., Bell M., eta Martin M. Learning Subjective Language. *Comput. Linguist.*, 30(3):277–308, September 2004. ISSN 0891-2017. URL <http://dx.doi.org/10.1162/0891201041850885>.
- Wiebe J.M., Bruce R.F., eta O'Hara T.P. Development and Use of a Gold-standard Data Set for Subjectivity Classifications. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, 246–253, Stroudsburg, PA, USA, 1999. Association for Computational Linguistics. ISBN 1-55860-609-3. URL <https://doi.org/10.3115/1034678.1034721>.

BIBLIOGRAFIA

- Wiegand M., Balahur A., Roth B., Klakow D., eta Montoyo A. A Survey on the Role of Negation in Sentiment Analysis. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, NeSp-NLP '10*, 60–68, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1858959.1858970>.
- Wilson T.A. *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. Doktoretza-tesia, University of Pittsburgh, 2008.
- Yu H. eta Hatzivassiloglou V. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP '03*, 129–136, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. URL <https://doi.org/10.3115/1119355.1119372>.
- Zhang W., Yu C., eta Meng W. Opinion Retrieval from Blogs. *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, CIKM '07*, 831–840, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-803-9. URL <http://doi.acm.org/10.1145/1321440.1321555>.

Terminologia eta laburdurak

Sentimenduen analisi (*Sentiment analysis*)

Orientazio semantiko (*Semantic orientation*)

Sentimendu-balentzi (*Sentiment valence*)

(Testuinguruko) balentzia-aldatzaile (*(Contextual) valence shifter*)

Sentimenduen sailkatzaile (*Sentiment classifier*)

Sentimenduen lexikoi (*Sentiment lexicon*)

Murritzapen Gramatika, MG (*Constraint Grammar, CG*)

Ezeztapen-marka (*Negation mark*)

(Ezeztapenaren) irismena (*Scope*)

Egitura lexikalizatu (*Lexicalized structure*)

Egitura Erretorikoaren Teoria (*Rhetorical Structure Theory, RST*)

Unitate zentrala, UZ (*Central Unit, CU*)

Erlaziozko diskurtso-egitura (*Relational discourse structure*)

BIBLIOGRAFIA

ERANSKINA

Murriztapen Gramatikako erregelak

Eranskin honetan ezeztapen-markak eta beren irismena identifikatzeko erregelak aurkezten ditugu. Erregelak Murriztapen Gramatika (MG) hurbilpean idatzirik daude.

```
LIST PUNTUAZIOA = PUNTPUNT PUNT_KOMA PUNT_BLPUNT  
PUNT_GALD PUNT_ESKL PUNT_HIRU PUNT_PUNT_KOMA;
```

```
LIST EZ = ‘ez’;  
LIST EZA = ‘ez’;
```

```
LIST BAINO = ‘baino’;  
LIST BESTERIK = ‘beste’;  
LIST ZALANTZARIK = ‘zalantza’;  
LIST DUDARIK = ‘duda’;  
LIST BESTERIKGABE = ‘besterik_gabe’;  
LIST EZINEZKOA = ‘ezinezko’;  
LIST EZINEAN = ‘ezinean’ ”ezin”;  
LIST NORA = ‘nora’;  
LIST EZEAN = ‘ezean’;
```

```

LIST NORAEZEAN = ‘‘nora_ezean’’;

LIST EZINIZAN = ‘‘ezin_izan’’;
LIST EZIN = ‘‘ezin’’ ‘‘ezindu’’;
LIST EZINHOBEA = ‘‘ezin_hobe’’;
LIST EZINIK = ‘‘ezinik’’;

LIST SALBU = ‘‘salbu’’;
LIST EZEZIK = ‘‘ez_ezik’’;
LIST IZANEZIK = ‘‘izan_ezik’’;

LIST GABE = ‘‘gabe’’;

LIST JUNTAGAILUA = ‘‘edo’’ ‘‘ala’’ ‘‘edota’’;
LIST BAI = ‘‘bai’’;

LIST EZTA = ‘‘ezta’’;

#MAPPINGS

MAP (!gabe) TARGET (ADB) IF (0C GABE);
MAP (!ez) TARGET (PRT) IF (0C EZ);
MAP (!eza) TARGET (IZE) IF (0C EZA);

#LEXIKALIZAZIOAK

# (1)
# ‘‘baino ez’’ egitura
MAP (!bainoezHAS) TARGET (LOT) IF (0C BAINO) (1C EZ);

# (2)
# ‘‘besterik ez’’ egiturak
MAP (!besterikezHAS) TARGET (DET) IF (0C BESTERIK) (1
  C EZ);

```

```

#MAP (!besterikezBUK) TARGET (PRT) IF (-1C BESTERIK)
    (0C EZ);

# (3)
# ‘zalantzarik gabe’
MAP (!zalantzarikgabeHAS) TARGET (IZE) IF (0C
    ZALANTZARIK) (1C GABE);
#MAP (!zalantzarikgabeBUK) TARGET (ADB) IF (-1C
    ZALANTZARIK) (0C GABE);

# (4)
# ‘dudarik gabe’
MAP (!dudarikgabeHAS) TARGET (IZE) IF (0C DUDARIK) (1C
    GABE);
#MAP (!dudarikgabeBUK) TARGET (ADB) IF (-1C DUDARIK) (0
    C GABE);

# (5)
# ‘besterik gabe’ antzemateko
MAP (!besterikgabe) TARGET (LOT) IF (0C BESTERIKGABE);

# (6)
# ‘ezinezkoa’ antzemateko
MAP (!ezinezkoa) TARGET (ADJ) IF (0 EZINEZKOA);

#(7)
# ezinean antzemateko
MAP (!ezinean) TARGET (ADB) IF (0C EZINEAN);

#(8)
# ‘nora ezean’ egitura.
MAP (!noraezeanHAS) TARGET (ADB) IF (0C NORA) (1C EZEAN
    );
MAP (!noraezeanBUK) TARGET (ADB) IF (-1C NORA) (0C

```

```

EZEAN);
MAP (!noraezean) TARGET (ADB) IF (0C NORAEZEAN);

##ERREGELA ZEHATZAK

#(1)
# EZIN: ‘‘ezin izan’’
MAP (!ezinizanHAS) TARGET (ADI) IF (0C EZINIZAN);
MAP (!ezinizanTAR1) TARGET (ADL) OR (ADI) OR (ADT) OR (
    IZE) OR (ADJ) OR (DET) OR (IOR) OR (LOT) IF (-1
    EZINIZAN);
MAP (!ezinizanTAR2) TARGET (ADI) OR (ADL) OR (IZE) OR (
    ADJ) OR (DET) OR (IOR) OR (LOT) IF (-2 EZINIZAN);
#MAP (!ezinizanBUK1) TARGET (ADI) OR (IZE) IF (-2
    EZINIZAN);
MAP (!ezinizanBUK2) TARGET (ADI) OR (ADL) OR (IZE) OR (
    ADJ) OR (DET) OR (IOR) OR (LOT) IF (-3 EZINIZAN) (
    NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!ezinizanBUK3) TARGET (ADI) OR (ADL) OR (IZE) OR (
    ADJ) OR (DET) OR (IOR) OR (LOT) IF (-4 EZINEAN) (NOT
    -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!ezinizanBUK4) TARGET (ADI) OR (ADL) OR (IZE) OR (
    ADJ) OR (DET) OR (IOR) OR (LOT) IF (-5 EZINIZAN) (
    NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2
    PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!ezinizanBUK5) TARGET (ADI) OR (ADL) OR (IZE) OR (
    ADJ) OR (DET) OR (IOR) OR (LOT) IF (-6 EZINIZAN) (
    NOT -5 PUNTUAZIOA) (NOT -4 PUNTUAZIOA) (NOT -3
    PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!ezinizanBUK6) TARGET (ADI) OR (ADL) OR (IZE) OR (
    ADJ) OR (DET) OR (IOR) OR (LOT) IF (-7 EZINIZAN) (
    NOT -6 PUNTUAZIOA) (NOT -5 PUNTUAZIOA) (NOT -4
    PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA)
    (NOT -1 PUNTUAZIOA);

```

MAP (!ezinizanBUK7) TARGET (ADI) OR (ADL) OR (IZE) OR (ADJ) OR (DET) OR (IOR) OR (LOT) IF (-8 EZINIZAN) (NOT -7 PUNTUAZIOA) (NOT -6 PUNTUAZIOA) (NOT -5 PUNTUAZIOA) (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

EZIN orokorra

MAP (!EZIN) TARGET (IZE) IF (0C EZIN);

MAP (!ezinik) TARGET (ADB) IF (0C EZINIK);

EZIN: ‘‘ezinik’’

MAP (!ezinik2) TARGET (ADI) IF (1C EZINIK);

MAP (!ezinik1) TARGET (IZE) IF (2C EZINIK);

###KOPONduta

EZIN: ‘‘ezin’’ + beste aditzak

MAP (!ezinTAR) TARGET (ADT) OR (ADI) OR (ADL) OR (ADLIZEELI) IF (-1C EZIN);

MAP (!ezinTAR2) TARGET (ADI) OR (ADL) OR (ADLIZEELI) IF (-2C EZIN);

MAP (!ezinBUK) TARGET (ADI) OR (ADL) OR (ADLIZEELI) IF (-3C EZIN);

EZIN: ‘‘ezin hobia’’

MAP (!ezinhobia) TARGET (ADJ) IF (0C EZINHOBIA);

#MAP (!ezinadjHAS) TARGET (IZE) OR (ADB) IF (0C EZIN);

MAP (!ezinadjBUK) TARGET (ADJ) OR (ADB) IF (-1C EZIN);

SALBU

MAP (!salbuBUK) TARGET (ADB) IF (0C SALBU);

MAP (!salbuHAS1) TARGET (IZE) OR (ADT) IF (1C SALBU);

MAP (!salbuHAS3) TARGET (IZE) OR (DET) IF (2C SALBU);

```

MAP (!salbuHAS3) TARGET (IZE) IF (3C SALBU);

# IZAN EZIK
MAP (!izanezik) TARGET (LOT) IF (0C IZANEZIK);
MAP (!izanezikHAS2) TARGET (IZE) IF (4C IZANEZIK);
MAP (!izanezikHAS1) TARGET (DET) IF (3C IZANEZIK);
MAP (!izanezikTAR1) TARGET (IZE) OR (DET) IF (2C
    IZANEZIK);
MAP (!izanezikTAR2) TARGET (ADI) OR (IZE) (1C IZANEZIK)
    ;

# EZ EZIK
MAP (!ezezik) TARGET (LOT) IF (0C EZEZIK);

#(2)
# EDO/EDOTA/ALA EZ
MAP (!juntagailuaHAS1) TARGET (ADI) OR (ADL) OR (ADT)
    OR (PRT) IF (1C JUNTAGAILUA) (2C EZ);
MAP (!juntagailuaHAS2) TARGET (ADI) OR (ADL) OR (IOR)
    IF (2C JUNTAGAILUA) (3C EZ);
MAP (!juntagailuaHAS3) TARGET (ADI) OR (IZE) IF (3C
    JUNTAGAILUA) (4C EZ);
MAP (!juntagailuaTAR) TARGET (LOT) IF (0C JUNTAGAILUA)
    (1C EZ);
#MAP (!juntagailuaBUK) TARGET (PRT) IF (-1C JUNTAGAILUA
    ) (0C EZ);

# EZ... EZ...
MAP (!ezez0) TARGET (PRT) IF (0C EZ) (2C EZ);
MAP (!ezez1) TARGET (DET) OR (IZE) OR (ADJ) IF (-1C EZ)
    (1C EZ);
MAP (!ezez2) TARGET (PRT) IF (-2C EZ) (0C EZ);
MAP (!ezez3) TARGET (DET) OR (IZE) OR (ADJ) IF (-1C EZ)
    (-3C EZ);

```

```

# EZ egiturak
MAP (!EZ-3) TARGET (IZE) IF (3C EZ) (NOT 2 BAINO) (NOT
  2 BESTERIK) (NOT 1 PUNTUAZIOA) (NOT 2 PUNTUAZIOA);
MAP (!EZ-2) TARGET (IZE) OR (ADI) OR (ADB) IF (2C EZ) (
  NOT 1C BAINO) (NOT 1C BESTERIK) (NOT 1C PUNTUAZIOA);
MAP (!EZ-1) TARGET (ADI) OR (ADJ) OR (ADL) OR (ADB) IF
  (1C EZ);
#MAP (!EZ0) TARGET (PRT) IF (0C EZ);
MAP (!EZ1) TARGET (ADT) OR (ADL) OR (ADI) OR (ADB) OR (
  IZE) OR (DET) OR (IOR) IF (-1C EZ) (NOT -2 BAINO) (
  NOT -2 BESTERIK);
MAP (!EZ2) TARGET (ADI) OR (ADL) OR (ADT) OR (DET) OR (
  IZE) OR (ADJ) OR (ADB) OR (IOR) IF (-2C EZ) (NOT -3C
  BAINO) (NOT -3C BESTERIK) (NOT -2 PUNTUAZIOA) (NOT
  -1 PUNTUAZIOA);
MAP (!EZ3) TARGET (ADI) OR (ADL) OR (ADT) OR (IZE) OR (
  ADJ) OR (IOR) IF (-3C EZ) (NOT -4C BAINO) (NOT -4
  BESTERIK) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (
  NOT -1 PUNTUAZIOA);
MAP (!EZ4) TARGET (ADI) OR (ADL) OR (IZE) OR (DET) OR (
  IOR) IF (-4C EZ) (NOT -5C BAINO) (NOT -5 BESTERIK) (
  NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2
  PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!EZ5) TARGET (ADI) OR (IZE) OR (IOR) OR (ADT) OR (
  ADL) IF (-5C EZ) (NOT -6C BAINO) (NOT -6 BESTERIK) (
  NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2
  PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!EZ6) TARGET (ADI) OR (IZE) OR (IOR) IF (-6C EZ) (
  NOT -7C BAINO) (NOT -7 BESTERIK) (NOT -5 PUNTUAZIOA)
  (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2
  PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!EZ7) TARGET (ADI) OR (IZE) OR (IOR) IF (-7C EZ) (
  NOT -8C BAINO) (NOT -8 BESTERIK) (NOT -6 PUNTUAZIOA)

```

```

        (NOT -5 PUNTUAZIOA) (NOT -4 PUNTUAZIOA) (NOT -3
        PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!EZ8) TARGET (ADI) OR (IZE) OR (ADL) OR (IOR) IF
        (-8C EZ) (NOT -9C BAINO) (NOT -9 BESTERIK) (NOT -7
        PUNTUAZIOA) (NOT -6 PUNTUAZIOA) (NOT -5 PUNTUAZIOA)
        (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2
        PUNTUAZIOA) (NOT -1 PUNTUAZIOA);
MAP (!EZ9) TARGET (ADI) IF (-9C EZ) (NOT -10C BAINO) (
        NOT -10 BESTERIK) (NOT -8 PUNTUAZIOA) (NOT -7
        PUNTUAZIOA) (NOT -6 PUNTUAZIOA) (NOT -5 PUNTUAZIOA)
        (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2
        PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

MAP (!izeezaHAS) TARGET (IZE) IF (1C EZA);
MAP (!izenezaBUK) TARGET (IZE) IF (0C EZA);

# GABE egiturak

MAP (!gabeAUR) TARGET (IZE) OR (ADI) OR (ADJ) OR (IOR)
        IF (1C GABE);
MAP (!gabeAUR2) TARGET (IZE) OR (ADI) IF (2C GABE);
MAP (!gabeAUR3) TARGET (IZE) IF (3C GABE);

# EZTA
#MAP (!EZTA0) TARGET (LOT) IF (0C EZTA);
MAP (!EZTA1) TARGET (ADT) OR (ADL) OR (ADI) OR (ADB) OR
        (IZE) OR (DET) OR (IOR) IF (-1C EZTA) (NOT -2 BAINO
        ) (NOT -2 BESTERIK);
MAP (!EZTA2) TARGET (ADI) OR (ADL) OR (ADT) OR (DET) OR
        (IZE) OR (ADJ) OR (ADB) OR (IOR) IF (-2C EZTA) (NOT
        -3C BAINO) (NOT -3C BESTERIK) (NOT -2 PUNTUAZIOA) (
        NOT -1 PUNTUAZIOA);
MAP (!EZTA3) TARGET (ADI) OR (ADL) OR (ADT) OR (IZE) OR

```

(ADJ) OR (IOR) IF (-3C EZTA) (NOT -4C BAINO) (NOT -4 BESTERIK) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

MAP (!EZTA4) TARGET (ADI) OR (ADL) OR (IZE) OR (DET) OR (IOR) IF (-4C EZTA) (NOT -5C BAINO) (NOT -5 BESTERIK) (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

MAP (!EZTA5) TARGET (ADI) OR (IZE) OR (IOR) OR (ADT) OR (ADL) IF (-5C EZTA) (NOT -6C BAINO) (NOT -6 BESTERIK) (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

MAP (!EZTA6) TARGET (ADI) OR (IZE) OR (IOR) IF (-6C EZTA) (NOT -7C BAINO) (NOT -7 BESTERIK) (NOT -5 PUNTUAZIOA) (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

MAP (!EZTA7) TARGET (ADI) OR (IZE) OR (IOR) IF (-7C EZTA) (NOT -8C BAINO) (NOT -8 BESTERIK) (NOT -6 PUNTUAZIOA) (NOT -5 PUNTUAZIOA) (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

MAP (!EZTA8) TARGET (ADI) OR (IZE) OR (ADL) OR (IOR) IF (-8C EZTA) (NOT -9C BAINO) (NOT -9 BESTERIK) (NOT -7 PUNTUAZIOA) (NOT -6 PUNTUAZIOA) (NOT -5 PUNTUAZIOA) (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

MAP (!EZTA9) TARGET (ADI) IF (-9C EZTA) (NOT -10C BAINO) (NOT -10 BESTERIK) (NOT -8 PUNTUAZIOA) (NOT -7 PUNTUAZIOA) (NOT -6 PUNTUAZIOA) (NOT -5 PUNTUAZIOA) (NOT -4 PUNTUAZIOA) (NOT -3 PUNTUAZIOA) (NOT -2 PUNTUAZIOA) (NOT -1 PUNTUAZIOA);

Tesi honen idazketa
2019ko urriaren 15ean
bukatu zen.